

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 00:30:12 ; Search time 66.4937 Seconds
(without alignments)
10792.671 Million cell updates/sec

Title: US-09-698-295-10
Perfect score: 14333
Sequence: 1 MWSEEEEDGDAETQDSE.....KLGPKASRSHNNKLOSTAS 2781

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.vrirus:*
- 16: sp.bacteriaph:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14333	100.0	2781	4 Q9UIG2	Q9Uig2 homo sapien
2	3535	24.7	796	11 Q8CFX5	Q8cfx5 mus musculu
3	3445	24.0	803	11 Q8K092	Q8k092 mus musculu
4	2848.5	19.9	2649	5 Q9W0T2	Q9w0t2 drosophila
5	2842.5	19.8	2669	5 Q9W0T1	Q9w0t1 drosophila
6	2839.5	19.8	2669	5 Q95VB8	Q95vb8 drosophila
7	2775.5	19.4	645	11 Q8VDN7	Q8vbn7 mus musculu
8	1368.5	9.5	412	4 Q9H5E0	Q9h5e0 homo sapien
9	871.5	6.1	1711	5 Q45409	Q45409 caenorhabdi
10	865.5	6.0	1713	5 Q95Z08	Q95zu8 caenorhabdi
11	674.5	4.7	1022	5 Q960Y3	Q960y3 drosophila
12	580.5	4.1	7210	5 Q9V7G8	Q9v7g8 drosophila
13	580.5	4.1	9270	5 Q8MLD9	Q8mld9 drosophila
14	572.5	4.0	2768	5 Q9VC00	Q9vc00 drosophila
15	572	4.0	17352	5 Q95YK2	Q95ym2 procambarus
16	560	3.9	5476	5 Q9NJ17	Q9nj17 drosophila

ALIGNMENTS

RESULT 1

Q9UIG2 PRELIMINARY; PRT; 2781 AA.
AC Q9UIG2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bromodomain PHD finger transcription factor.
GN BPTF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=20130111; PubMed=10662542;
RX Jones M.H., Hamana N., Shimane M.;
RT "Identification and characterization BPTF, a novel bromodomain
transcription factor";
RL Genomics 63:35-39(2000).
DR EMBL; AB032251; BAA89208.1;
DR HSP; Q92831; 1B91.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02791; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 2.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2781 AA; 311210 MW; 7B3BE0C1C4C3CFBE CRC64;

Query Match

100.0%; Score 14333; DB 4; Length 2781;

Best Local Similarity 100.0%; Pred. No. 0;		Matches 2781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MYSEEEEDGDAEETQSEDEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST	60
Db	1	MYSEEEEDGDAEETQSEDEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST	60
Qy	61	YSSTPGARKVPRPRPILPEKIDPLEPFPKSSDLMVNEHIMNIAIYEVLRNFTV	120
Db	61	YSSTPGARKVPRPRPILPEKIDPLEPFPKSSDLMVNEHIMNIAIYEVLRNFTV	120
Qy	121	LRLSPFRFDFCAALVSOEQCTLAEMHVLLKAVLREEDSNTTFPADLKDSVNSTLY	180
Db	121	LRLSPFRFDFCAALVSOEQCTLAEMHVLLKAVLREEDSNTTFPADLKDSVNSTLY	180
Qy	181	FIDGWTWPEVLRYVCSKXYHHVLPYQEAEDYPGVNKKIKVLQFLVDQFLTNIARE	240
Db	181	FIDGWTWPEVLRYVCSKXYHHVLPYQEAEDYPGVNKKIKVLQFLVDQFLTNIARE	240
Qy	241	ELMSEGVIOYDDHCRVCHKLDLCCETCSAVYHLECVKPPLEVPDEMOCEVCVAHKV	300
Db	241	ELMSEGVIOYDDHCRVCHKLDLCCETCSAVYHLECVKPPLEVPDEMOCEVCVAHKV	300
Qy	301	PGVTDCAEIQKKNPYIRHEPIGYDRSRKRYWFLNRLLIIBEDTENENEKKIWTYSTKVQ	360
Db	301	PGVTDCAEIQKKNPYIRHEPIGYDRSRKRYWFLNRLLIIBEDTENENEKKIWTYSTKVQ	360
Qy	361	LAELIDCLDKDYWAELCKLILEENREETHRHMDITDITNKARGSNKSFLLAANEEILES	420
Db	361	LAELIDCLDKDYWAELCKLILEENREETHRHMDITDITNKARGSNKSFLLAANEEILES	420
Qy	421	IRAKKGIDINVKSPEETEKDNKNETENDSKAERREEFEDQSLKSDDKTDPDDPEQK	480
Db	421	IRAKKGIDINVKSPEETEKDNKNETENDSKAERREEFEDQSLKSDDKTDPDDPEQK	480
Qy	481	SEVGDFKSEKSNGLSESPGAGKAGCAGSTRITILRNPDSKLSOLKSOVAAAHAENKL	540
Db	481	SEVGDFKSEKSNGLSESPGAGKAGCAGSTRITILRNPDSKLSOLKSOVAAAHAENKL	540
Qy	541	FKEKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLQEGKYRVYHNOYSTNFALNKH	600
Db	541	FKEKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLQEGKYRVYHNOYSTNFALNKH	600
Qy	601	QHRDHDKRRLAHKFLTPAGEFKWGSVHGSKVLITSLRLITITOLENNIPSFLLHPN	660
Db	601	QHRDHDKRRLAHKFLTPAGEFKWGSVHGSKVLITSLRLITITOLENNIPSFLLHPN	660
Qy	661	WASHRANWIKAVQMSKPREFALALILECAVKPVWMLPIWREPLGHTRLHMTSIEREE	720
Db	661	WASHRANWIKAVQMSKPREFALALILECAVKPVWMLPIWREPLGHTRLHMTSIEREE	720
Qy	721	KEYVKKKEKKOEETQQAATWVKYTPPVKHQVWKQGEERYVTGYGWSWISKTHVYRF	780
Db	721	KEYVKKKEKKOEETQQAATWVKYTPPVKHQVWKQGEERYVTGYGWSWISKTHVYRF	780
Qy	781	VPLKPGNTNRYKSLSTKNNMDENDESCKRSRPPKIKIETPSEKDEVKGSDAK	840
Db	781	VPLKPGNTNRYKSLSTKNNMDENDESCKRSRPPKIKIETPSEKDEVKGSDAK	840
Qy	841	GADQNMIDISKITEKKQDVYKELLDSDDKPEEPMEVDMDKTESHVNCQESSQVDVY	900
Db	841	GADQNMIDISKITEKKQDVYKELLDSDDKPEEPMEVDMDKTESHVNCQESSQVDVY	900
Qy	901	NVSGFHLRTSYKKKTKSSKLDGLLERRIKOFTLEEKORLEKIKLEGGIKIGTSTNSS	960
Db	901	NVSGFHLRTSYKKKTKSSKLDGLLERRIKOFTLEEKORLEKIKLEGGIKIGTSTNSS	960
Qy	961	KNLSSEPVITAKEGGQSDSMRQSQSNANNDQEDLIQGCQSDSSVLRMSDPSHTNK	1020
Db	961	KNLSSEPVITAKEGGQSDSMRQSQSNANNDQEDLIQGCQSDSSVLRMSDPSHTNK	1020
Qy	1021	LYPKDRVLDVYSIRSPETKCPKQNSIENDIEEKVSDLASRGOEPTKSKTKGNDFFIDDSK	1080
Db	1021	LYPKDRVLDVYSIRSPETKCPKQNSIENDIEEKVSDLASRGOEPTKSKTKGNDFFIDDSK	1080
Qy	1081	LASADDICTLCKNKKPLIQUESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMDFEGL	1140
Db	1081	LASADDICTLCKNKKPLIQUESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMDFEGL	1140
Qy	1141	GCDSESNSTLNSDTSVSIQDSSEEDMIVQNSNESISEQFRTREODVLEPLKCELVS	1200
Db	1141	GCDSESNSTLNSDTSVSIQDSSEEDMIVQNSNESISEQFRTREODVLEPLKCELVS	1200
Qy	1201	ESTGNCEDRLPVKGTPEANGKPKSQOKLEERPVNKCSDQIKLNTTDDKNNENRESEKKG	1260
Db	1201	ESTGNCEDRLPVKGTPEANGKPKSQOKLEERPVNKCSDQIKLNTTDDKNNENRESEKKG	1260
Qy	1261	QRTSTFQINGDKNPKIYLKGECLKEISESVVSVNVEPKVNNINKIIPENDIKSLTVKE	1320
Db	1261	QRTSTFQINGDKNPKIYLKGECLKEISESVVSVNVEPKVNNINKIIPENDIKSLTVKE	1320
Qy	1321	SAIRPFINGDVIMEDFERNESSETKSHLLSSDABGNRYRDSLETLPSPKESDSTOTTPS	1380
Db	1321	SAIRPFINGDVIMEDFERNESSETKSHLLSSDABGNRYRDSLETLPSPKESDSTOTTPS	1380
Qy	1381	ASCPESNSVQVEDMEIETSEVKKVTSSPITSEESNLSNDFIDENGLPINKNENNGES	1440
Db	1381	ASCPESNSVQVEDMEIETSEVKKVTSSPITSEESNLSNDFIDENGLPINKNENNGES	1440
Qy	1441	KRKTVITETVTMTSTVATESKTIVKVEKDKQTVVSVSTENCACKSTVTTTTTWTKLSTPS	1500
Db	1441	KRKTVITETVTMTSTVATESKTIVKVEKDKQTVVSVSTENCACKSTVTTTTTWTKLSTPS	1500
Qy	1501	TGGSVDIISVKEQSKTVVTVTDSLTGTTGTLVTSMTVSKVSTYTRDKVKLMKFSRPKKT	1560
Db	1501	TGGSVDIISVKEQSKTVVTVTDSLTGTTGTLVTSMTVSKVSTYTRDKVKLMKFSRPKKT	1560
Qy	1561	RSGTALPSYRKFVYKSTKTSIFVLNDDLLKLARKGGIREVYFVYNNAKPALDIMPYPSP	1620
Db	1561	RSGTALPSYRKFVYKSTKTSIFVLNDDLLKLARKGGIREVYFVYNNAKPALDIMPYPSP	1620
Qy	1621	RPTFGITWRYLQVTVKSLAGVSLMLRLWASLRWDDMAKVPVGGSGTSTETSEITITT	1680
Db	1621	RPTFGITWRYLQVTVKSLAGVSLMLRLWASLRWDDMAKVPVGGSGTSTETSEITITT	1680
Qy	1681	EIIKRRDVGPIREYCIIRKIIICIGVPEPKETPTORGLRSALSARPKRPETPKOTG	1740
Db	1681	EIIKRRDVGPIREYCIIRKIIICIGVPEPKETPTORGLRSALSARPKRPETPKOTG	1740
Qy	1741	PVIETWAELELEWEIRAEFAERVEKEKAQAVEQAKRLEQKPTVIATSTTST	1800
Db	1741	PVIETWAELELEWEIRAEFAERVEKEKAQAVEQAKRLEQKPTVIATSTTST	1800
Qy	1801	TSTISPAOKVWAPISGSVTTGTMVLTTKVGSPTATVFOONKNPHOTFATVWKGQNS	1860
Db	1801	TSTISPAOKVWAPISGSVTTGTMVLTTKVGSPTATVFOONKNPHOTFATVWKGQNS	1860
Qy	1861	GWVQVQKVLGIIPSTSTSOOTFTSFOPRATVIRPNTSGSGGTTNSOVITGPOLRP	1920
Db	1861	GWVQVQKVLGIIPSTSTSOOTFTSFOPRATVIRPNTSGSGGTTNSOVITGPOLRP	1920
Qy	1921	GNTVIRTPLOQSTLGKALIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNTVSTPGOK	1980
Db	1921	GNTVIRTPLOQSTLGKALIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNTVSTPGOK	1980
Qy	1981	SLTSATSTNSIQSSASQPPRQOQGVKLTMAQLTOLTQGHGNGQLTVVIOGGQTTGOL	2040
Db	1981	SLTSATSTNSIQSSASQPPRQOQGVKLTMAQLTOLTQGHGNGQLTVVIOGGQTTGOL	2040
Qy	2041	QLIPGVTVLPGCQQLMQAAMPNCTVOREFLTPLATTATTASTTTTTVSTTAAGTGBOR	2100
Db	2041	QLIPGVTVLPGCQQLMQAAMPNCTVOREFLTPLATTATTASTTTTTVSTTAAGTGBOR	2100
Qy	2101	QSKLSQPMQVHODKTLTPPAQSSSVGPAKAQPTQAPSAARPQOTQPSQPAQPEVOTQPEV	2160
Db	2101	QSKLSQPMQVHODKTLTPPAQSSSVGPAKAQPTQAPSAARPQOTQPSQPAQPEVOTQPEV	2160

KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 803 AA; 85647 MW; 731020601D158142 CRC64;
Query Match 24.0%; Score 3445; DB 11; Length 803;
Best Local Similarity 84.6%; Pred. No. 6.6e-140;
Matches 680; Conservative 44; Mismatches 68; Indels 12; Gaps 3;
QY 1378 TPSASCPESNVNOVEDMEIETSEVKKVTSPTSEESNLSDNFDIENGLPINKNNVN 1437
DQ 5 TPRAPCPDRNSLQVEDMETESPESVKRVPSPVTGEGSNLSKGFMDNGLPSSKDNVN 64
QY 1438 GESKRTKIVETVMTSTVATESKTVIKVEKDKQTQVVSSTENCAKSTVTTTTTIVTKLS 1497
DQ 65 GESKRTKIVETVMTSTVATESKTVIKVAKGDKQTQVVSSTENCAKSTVTTTTTIVTKLS 124
QY 1498 TSPGSGVDIISVKEQSKTVTTTVDLSLTGTGLTMTSVTSKEYSTRDKVKLMKFSRP 1557
DQ 125 TSPDPTGVDIISVKEQSKTVTTTVDLSLTGTGLTMTSVTSKEYSTRDKVKLMKFSRP 184
QY 1558 KTRSGTALPSYKRFVTKSKTSIFVLPNDLKLARKGGIREVPPYNYNAKPAIDWIPY 1617
DQ 185 KTRSGTALPSYKRFVTKSKTSIFVLPNDLKLARKGGIREVPPYNYNAKPAIDWIPY 244
QY 1618 PSPRPTGIIWRYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGGSGSTRTSTSEI 1677
DQ 245 PSPRPTGIIWRYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGGSGSTRTSTSEI 304
QY 1678 TTTEIIRKRDVPGYIGIRFPCYKIRKIIICPGVPEPKPTPPQRKGLRSSALRPKRPETPK 1737
DQ 305 TTTEIIRKRDVPGYIGIRFPCYKIRKIIICPGVPEPKPTPPQRKGLRSSALRPKRPETPK 364
QY 1738 QTGPVITETWAEELWEIRAFERAEVEKAKAQVQQAQKRLQEQKFPVIASTTSPT 1797
DQ 365 QTGPVITETWAEELWEIRAFERAEVEKAKAQVQQAQKRLQEQKFPVIASTTSPT 424
QY 1798 SSTTSTSPAKVMVAPISGVTGKVLTKVGSPTATVTFQONKFNHOFATFWKQGG 1857
DQ 425 SSTTSTSPAKVMVAPISGVTGKVLTKVGSPTATVTFQONKFNHOFATFWKQGG 484
QY 1858 SNSGVVOQKVLGIIPSSGTSTQSTSTQSTSTQSTSTQSTSTQSTSTQSTSTQST 1917
DQ 485 SNSGVVOQKVLGIIPSSGTSTQSTSTQSTSTQSTSTQSTSTQSTSTQSTSTQST 544
QY 1918 IRGEMTVIRPLQOSTLGKILIRPVMVQPCAPQVNTQIIRGQPVSTAYSAPNTVSTP 1977
DQ 545 IRGEMTVIRPLQOSTLGKILIRPVMVQPCAPQVNTQIIRGQPVSTAYSAPNTVSTP 604
QY 1978 GQSLTSATSTSNIOSSASOPPRPOQGVKLTMAQLTQLTGCGHGNGLTVVIOGQGT 2037
DQ 505 GQSLTSATSTSNIOSSASOPPRPOQGVKLTMAQLTQLTGCGHGNGLTVVIOGQGT 664
QY 2038 GQLQILPGVTVLPQGGQQLQAAMPNGTVQRFLETPLATATATAS ---TTTTTSTTA 2093
DQ 665 GQLQILPGVTVLPQGGQQLQAAMPNGTVQRFLETPLATATATAS ---TTTTTSTTA 724
QY 2094 AGTGEQKSLSPOMVHODKTLPPAQSSVGPAPKAPQAPQAPQAPQAPQAPQAPQ 2153
DQ 725 AGTGEQKSLSPOMVHODKTLPPAQSSVGPAPKAPQAPQAPQAPQAPQAPQAPQ 782
QY 2154 VQTEQV 2177
DQ 783 VQTEQV 800
RESULT 4
Q9W0T2 PRELIMINARY; PRT: 2649 AA.
AC Q9W0T2
DT 01-MAR-2000 (TrEMBLrel. 13, Created)
DF 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG32346 protein.

GN CG10894.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
KC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Helt G., Nelson C.R., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RN SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Pfeiffer B.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome.";
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

```
QY 2161 QTQTVSSHPSEAPTHAQSSKPVAAQSQPNVQSQSPVRVQSPQTRIRPSTPSQL 2220
DB 2161 QTQTVSSHPSEAPTHAQSSKPVAAQSQPNVQSQSPVRVQSPQTRIRPSTPSQL 2220
QY 2221 SPGQSQVQTTSPQPIQPIQTHSLOIPSGQGOPOQVQSSQTOFLSSQOTLNQVSVSSPS 2280
DB 2221 SPGQSQVQTTSPQPIQPIQTHSLOIPSGQGOPOQVQSSQTOFLSSQOTLNQVSVSSPS 2280
QY 2281 RPQIQIQPOQVITAVPOLQOQOVLSQIQSQVVAQIQAAQSGVPPQIKLQPLQIQOSS 2340
DB 2281 RPQIQIQPOQVITAVPOLQOQOVLSQIQSQVVAQIQAAQSGVPPQIKLQPLQIQOSS 2340
QY 2341 AVQTHQIQNVVYVQAASVQEQQLRVQQLRDQOQKKQOQIIEIKREHTIQAQSQSEIIQOK 2400
DB 2341 AVQTHQIQNVVYVQAASVQEQQLRVQQLRDQOQKKQOQIIEIKREHTIQAQSQSEIIQOK 2400
QY 2401 VVMKHNAVIEHLKOKKSMTPAEREENQRMIVCNQVMKYILDKIDKEEKQAAKKRREESV 2460
DB 2401 VVMKHNAVIEHLKOKKSMTPAEREENQRMIVCNQVMKYILDKIDKEEKQAAKKRREESV 2460
QY 2461 EOKRSKONATKLSALLFKHKLQRAELKRALDKDLQIEVQBELKRDLLKKEKDLQ 2520
DB 2461 EOKRSKONATKLSALLFKHKLQRAELKRALDKDLQIEVQBELKRDLLKKEKDLQ 2520
QY 2521 LAQATAVAACPVPVLPAPAPPPPPPPVQHTGLJSTPTLPVASQKRKREEEKDS 2580
DB 2521 LAQATAVAACPVPVLPAPAPPPPPPPVQHTGLJSTPTLPVASQKRKREEEKDS 2580
QY 2581 SSKSKKKMISTTSKTKKOTKLYCICKTPYDESKEYIGCDRCNWHGRCVGIQSEAE 2640
DB 2581 SSKSKKKMISTTSKTKKOTKLYCICKTPYDESKEYIGCDRCNWHGRCVGIQSEAE 2640
QY 2641 LIDYVCPQCOSTEDAMTVLTPLEKDYEGRLKRLVLSLQAHKMAWPLEPVPDNDADPY 2700
DB 2641 LIDYVCPQCOSTEDAMTVLTPLEKDYEGRLKRLVLSLQAHKMAWPLEPVPDNDADPY 2700
QY 2701 GVKEPMDLATMERVORRYEKELEFVADNKTIFDNCRYNPNDSFPYQCAEVLESFFV 2760
DB 2701 GVKEPMDLATMERVORRYEKELEFVADNKTIFDNCRYNPNDSFPYQCAEVLESFFV 2760
QY 2761 QKLGFKASRSHNNKLOSTAS 2781
DB 2761 QKLGFKASRSHNNKLOSTAS 2781

RESULT 2
Q8CFX5 PRELIMINARY; PRT; 796 AA.
AC Q8CFX5,
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DE Similar to fetal alzheimer antigen (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC037661; AAH37661.1; -
FT NON_TER 796
SQ SEQUENCE 796 AA; 90592 MW; 2C1BBE28009C82F3 CRC64;

Query Match 24.7%; Score 3535; DB 11; Length 796;
Best Local Similarity 84.3%; Pred. No. 9e-144;
Matches 671; Conservative 35; Mismatches 24; Indels 66; Gaps 4;

QY 1 MVSEDEEEEDGDAEETQDSEDEEDDMEEDDDDDSDYPEEMEDDDDDASYCTESSFRSH 58
DB 1 MVSEDEEEEDGDAEETQDSEDEEDDMEEDDDDDSDYPEEMEDDDDDASYCTESSFRSH 58
QY 59 STYSTPGRKPRVHRPRSPILKEKDIPLEPKSSEDLMPNEHIMNVIAIYEVLRNFG 118
DB 59 STYSTPGRKPRVHRPRSPILKEKDIPLEPKSSEDLMPNEHIMNVIAIYEVLRNFG 118
QY 61 STYSTPGRKPRVHRPRSPILKEKDIPLEPKSSEDLMPNEHIMNVIAIYEVVRNFG 120
DB 61 STYSTPGRKPRVHRPRSPILKEKDIPLEPKSSEDLMPNEHIMNVIAIYEVVRNFG 120
QY 119 TVLRUSPFRFEDCAALYSQEOCTLMAEMHVVLKAVJREEDTSNTTFGPADLKDSVNST 178
DB 119 TVLRUSPFRFEDCAALYSQEOCTLMAEMHVVLKAVJREEDTSNTTFGPADLKDSVNST 178
QY 121 NVLRUSPFCFEDCAALYSQEOCTLMAEMHVALLKAVJREEDTSNTTFGPADLKDSVNST 180
DB 121 NVLRUSPFCFEDCAALYSQEOCTLMAEMHVALLKAVJREEDTSNTTFGPADLKDSVNST 180
QY 179 LYFIDGWTWPEVLRVYCESDKKEYHHVLPYQEAEDYPYGEVNEKIKVLFQVLDQFLTTNIA 238
DB 179 LYFIDGWTWPEVLRVYCESDKKEYHHVLPYQEAEDYPYGEVNEKIKVLFQVLDQFLTTNIA 238
QY 181 LYFIDGWTWPEVLRVYCESDKKEYHHVLPYQEAEDYPYGEVNEKIKVLFQVLDQFLTTNIA 240
DB 181 LYFIDGWTWPEVLRVYCESDKKEYHHVLPYQEAEDYPYGEVNEKIKVLFQVLDQFLTTNIA 240
QY 239 REELMSEGVIOYDDHCRVCHKLGLDLCCEETSAVYHLECVKPPLEEVPEDEQVCEVVAH 298
DB 239 REELMSEGVIOYDDHCRVCHKLGLDLCCEETSAVYHLECVKPPLEEVPEDEQVCEVVAH 298
QY 241 REELMSEGVIOYDDHCRVCHKLGLDLCCEETSAVYHLECVKPPLEEVPEDEQVCEVVAH 300
DB 241 REELMSEGVIOYDDHCRVCHKLGLDLCCEETSAVYHLECVKPPLEEVPEDEQVCEVVAH 300
QY 299 KVPVGTDCVAEIQKNKPIIRHEPIGYDRSRKRYWFLNRLIITEEDTENENKKIWYSTK 358
DB 299 KVPVGTDCVAEIQKNKPIIRHEPIGYDRSRKRYWFLNRLIITEEDTENENKKIWYSTK 358
QY 301 KVPVGTDCVAEYQKNKPYRHEPIGYDRSRKRYWFLNRLIITEEDTENENKKVWYSTK 360
DB 301 KVPVGTDCVAEYQKNKPYRHEPIGYDRSRKRYWFLNRLIITEEDTENENKKVWYSTK 360
QY 359 VOLAELIDCLDKDYWEAELCKTILEMREIEIHRHMDITEDLTNKGSKNSFLAAANEEL 418
DB 359 VOLAELIDCLDKDYWEAELCKTILEMREIEIHRHMDITEDLTNKGSKNSFLAAANEEL 418
QY 361 VOLAELIDCLDKDYWEAELCKTILEMREIEIHRHMDITEDLTNKGSKNSFLAAANEEL 420
DB 361 VOLAELIDCLDKDYWEAELCKTILEMREIEIHRHMDITEDLTNKGSKNSFLAAANEEL 420
QY 419 ESIRAKKG-DIDNVKSPSEETKDKNETEND-SKDAENREEFDQSLKSDKDDKDDDD 476
DB 419 ESIRAKKG-DIDNVKSPSEETKDKNETEND-SKDAENREEFDQSLKSDKDDKDDDD 476
QY 421 DSLIRKRGEDIDCDQSPEDPKKHGEGENSSKDAEADPSADKDADSKGLEEP 480
DB 421 DSLIRKRGEDIDCDQSPEDPKKHGEGENSSKDAEADPSADKDADSKGLEEP 480
QY 477 EQGKSE----- 482
DB 477 EQGKSE----- 482
QY 481 GHGKPEEPTGYDGNKSNVPAFLGNTTINASPEETSPCDGRSPGCLSFTHDSSMAEKV 540
DB 481 GHGKPEEPTGYDGNKSNVPAFLGNTTINASPEETSPCDGRSPGCLSFTHDSSMAEKV 540
QY 483 -----VGFKSKSNGLSESPGAGKSGSTRITITRLNRPDSKLSQKSOVAAAA 534
DB 483 -----VGFKSKSNGLSESPGAGKSGSTRITITRLNRPDSKLSQKSOVAAAA 534
QY 541 ASELPPDPVPGDFRLEKSGEVSSEPGAGKSGSTRITITRLNRPDSKLSQKSOVAAAA 600
DB 541 ASELPPDPVPGDFRLEKSGEVSSEPGAGKSGSTRITITRLNRPDSKLSQKSOVAAAA 600
QY 535 HEANKLPKGEKVLVYNSQGEISRLSTKKEVIMKNNINNYFKLGQEGKYRVYHNOYSTNS 594
DB 535 HEANKLPKGEKVLVYNSQGEISRLSTKKEVIMKNNINNYFKLGQEGKYRVYHNOYSTNS 594
QY 601 HEANKLPKGEKVLVYNSQGEISRLSTKKEVIMKNNINNYFKLGQEGKYRVYHNOYSTNS 660
DB 601 HEANKLPKGEKVLVYNSQGEISRLSTKKEVIMKNNINNYFKLGQEGKYRVYHNOYSTNS 660
QY 595 FALANKHOHREDHDKRHLAKFCLTPAGEFKWNGSVHGSKVLITSTLRITITOLENNIPS 654
DB 595 FALANKHOHREDHDKRHLAKFCLTPAGEFKWNGSVHGSKVLITSTLRITITOLENNIPS 654
QY 661 FALANKHOHREDHDKRHLAKFCLTPAGEFKWNGSVHGSKVLITSTLRITITOLENNIPS 720
DB 661 FALANKHOHREDHDKRHLAKFCLTPAGEFKWNGSVHGSKVLITSTLRITITOLENNIPS 720
QY 655 SFLEPNWASHRANWIKAVQMCSPREFALALALECAVKPVYMLPIWREFLGHTRLHRT 714
DB 655 SFLEPNWASHRANWIKAVQMCSPREFALALALECAVKPVYMLPIWREFLGHTRLHRT 714
QY 721 SFLEPNWASHRANWIKAVQMCSPREFALALALECAVKPVYMLPIWREFLGHTRLHRT 780
DB 721 SFLEPNWASHRANWIKAVQMCSPREFALALALECAVKPVYMLPIWREFLGHTRLHRT 780
QY 715 SIEREEKKVKKKKK 730
DB 715 SIEREEKKVKKKKK 730
QY 781 SIEREEKKVKKKKK 796
DB 781 SIEREEKKVKKKKK 796

RESULT 3
Q8K092 PRELIMINARY; PRT; 803 AA.
AC Q8K092;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032289; AAH32289.1; -
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 01:17:18 ; Search time 109.881 Seconds
(without alignments)
3939.095 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

Sequence: 1 MVSEEEEDGDAETQDSE.....KLGFKASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 148893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	9.3	328	9	US-09-925-297-816
2	982.5	6.6	238	9	US-09-729-835-80
3	982.5	3.0	3664	15	US-10-177-293-423
4	443	3.0	2344	9	US-09-815-242-12713
5	438	2.9	3256	10	US-09-919-172-98
6	438	2.9	3256	11	US-09-919-039-21
7	432	2.9	2665	9	US-09-864-761-34248
8	402	2.7	4019	10	US-09-738-973-425
9	402	2.7	4019	10	US-09-854-133-425
10	402	2.7	4019	15	US-10-144-649A-425
11	401.5	2.7	1367	10	US-09-801-368-108
12	391.5	2.6	3899	15	US-10-171-311-4
13	391.5	2.6	3917	15	US-10-171-311-8
14	391	2.6	5179	9	US-09-922-217-1068
15	391	2.6	5179	10	US-09-833-263-1068

16	391	2.6	5179	14	US-10-025-380-1068	Sequence 1068, Ap
17	389.5	2.6	2478	9	US-09-815-242-5816	Sequence 5816, Ap
18	389.5	2.6	2478	9	US-09-815-242-12967	Sequence 12967, A
19	387.5	2.6	3907	15	US-10-171-311-2	Sequence 2, Appli
20	387.5	2.6	3925	15	US-10-171-311-6	Sequence 6, Appli
21	384	2.6	5935	15	US-10-243-243A-8	Sequence 8, Appli
22	381.5	2.5	5877	15	US-10-142-515-11	Sequence 11, Appl
23	379	2.5	2701	15	US-10-171-311-83	Sequence 83, Appl
24	373	2.5	2476	11	US-09-824-574-7	Sequence 7, Appli
25	370	2.5	2297	15	US-10-245-802-20	Sequence 20, Appl
26	362.5	2.4	2843	10	US-09-987-482-1	Sequence 1, Appli
27	361.5	2.4	2843	8	US-08-681-219-32	Sequence 32, Appl
28	361.5	2.4	2843	12	US-10-092-138-30	Sequence 30, Appl
29	354.5	2.4	1400	10	US-09-764-176-7	Sequence 7, Appli
30	344.5	2.3	2382	15	US-10-196-935A-2	Sequence 2, Appli
31	342	2.3	2368	9	US-09-815-242-5635	Sequence 5635, Ap
32	342	2.3	2368	9	US-09-815-242-12389	Sequence 12389, A
33	339	2.3	6281	9	US-09-815-242-12996	Sequence 12996, A
34	337	2.3	528	12	US-09-840-746-20	Sequence 20, Appl
35	331	2.2	1236	11	US-09-769-787-109	Sequence 109, App
36	329.5	2.2	5795	9	US-09-815-242-12610	Sequence 12610, A
37	328	2.2	1863	10	US-09-734-672-4	Sequence 4, Appli
38	328	2.2	1863	11	US-09-982-828-6	Sequence 6, Appli
39	324	2.2	1781	9	US-09-738-877-3	Sequence 3, Appli
40	324	2.2	1781	11	US-09-961-403-13	Sequence 13, Appl
41	322	2.2	1863	10	US-09-734-672-2	Sequence 2, Appli
42	322	2.2	1863	10	US-09-734-672-6	Sequence 6, Appli
43	322	2.2	1863	11	US-09-982-828-2	Sequence 2, Appli
44	322	2.2	1863	11	US-09-982-828-4	Sequence 4, Appli
45	322	2.2	1863	15	US-10-022-819-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-925-297-816
; Sequence 816, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 816
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (170)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (172)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (174)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (178)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (183)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (269)

OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (286)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-925-297-816

Query Match 9.3%; Score 1385; DB 9; Length 328;
Best Local Similarity 66.4%; Pred. No. 4.3e-63;
Matches 286; Conservative 0; Mismatches 17; Indels 128; Gaps 2;

406 NKSFLAAANEELISIRAKKGDIDNVKSP--EETEKDKNETENDSKDAEKNREPEQSL 463
12 NVSLSGTNEELISIRAKKGDIDNVKSP--EETEKDKNETENDSKDAEKNREPEQSL 71
464 EKSDDDKTPDDPEQKSEPTTEVGDGKNSVANLGDNTTATSETPSEGRSPVGCLS 523
72 EKSDDDKTPDDPEQKSE----- 90
524 ETPDSSNMAEKVASELPQDPPEPKNKTCESNTSATTTISIQPNLENSSSSELNSQSE 583
91 ----- 90
584 SAKAADDPENGERSHPTVSIQIEIVGDFTSKSTGELSESPGAGKAGSGSTRIITRLRN 643
91 -----VGDFFKSEKNGELSESPGAGKAGSGSTRIITRLRN 125
644 PDKLSQLKSQVAAAHAANKLFKEGKEVLYVNSQGEISRLSTKKEVIMKNNINPKL 703
126 PDKLSQLKSQVAAAHAANKLFKEGKEVLYVNSQGEISRLSTKKEVIMKNNINPKL 185
704 GQEGKYRVYHNOYSTNSFALNKHQREDHDKRHLAHEKFLTPAGFEKNGSVHGSKVLT 763
186 GQEGKYRVYHNOYSTNSFALNKHQREDHDKRHLAHEKFLTPAGFEKNGSVHGSKVLT 245
764 ISTRLTITOLENNIPSSFPNASHRANWIKAVQMSKPREFALALAILCAVAKPVVM 823
246 ISTRLTITOLENNIPSSFPNASHRANWIKAVQMSKPREFALALAILCAVAKPVVM 305
824 LPIWREPLIGHT 834
306 LPIWREPLIGHT 316

RESULT 2
US-09-729-835-80
Sequence 80, Application US/09729835
Patent No. US20010016647A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 29 Human Secreted Proteins
FILE REFERENCE: P2015P1
CURRENT APPLICATION NUMBER: US/09/729,835
CURRENT FILING DATE: 2000-12-06
PRIOR APPLICATION NUMBER: 09/257,179
PRIOR FILING DATE: 1999-02-25
PRIOR APPLICATION NUMBER: 60/056,270
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: 60/056,271
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: 60/056,247
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: 60/056,073
PRIOR FILING DATE: 1997-08-29
NUMBER OF SEQ ID NOS: 128
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 80
LENGTH: 238
TYPE: PRT
ORGANISM: Homo sapiens
US-09-729-835-80

Query Match 6.6%; Score 982.5; DB 9; Length 238;
Best Local Similarity 78.8%; Pred. No. 9.8e-43;

Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;
2689 PTLIPVASQKRREK-----DSSSKSKKKMISTTSKTKK-----DT 2727
3 PTRP-----KTPYDESKFYIGCDLCTNWHGECVIGTEKEAKKMDVVICNDCKRAQSGSSE 58
2728 KLYCICKTPYDESKFYIGCDRCQNWYHGRVGLQSEALIDYVCPQCOSTEDAMTVLT 2787
59 ELYCICKTPYDESKFYIGCDRCQNWYHGRVGLQSEALIDYVCPQCOSTEDAMTVLT 118
2788 PLTEKDYEGKLKRVLSQAHKMAWPFLEPVDNDAPDYGVIKPEMDLATMEERVORRY 2847
119 PLTEKDYEGKLKRVLSQAHKMAWPFLEPVDNDAPDYGVIKPEMDLATMEERVORRY 178
2848 EKLTEFVADMTKIFDNCRYNPNPSPFYQCAEVLESEFFVQKLKGFKASRSHNNKLOSTAS 2907
179 EKLTEFVADMTKIFDNCRYNPNPSPFYQCAEVLESEFFVQKLKGFKASRSHNNKLOSTAS 238

RESULT 3
US-10-177-293-423
Sequence 423, Application US/10177293
Publication No. US20030124128A1
GENERAL INFORMATION:
APPLICANT: Lillie, Karen
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegu
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 423
LENGTH: 3664
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-423

Query Match 3.0%; Score 456.5; DB 15; Length 3664;
Best Local Similarity 18.9%; Pred. No. 2.6e-14;
Matches 571; Conservative 403; Mismatches 1098; Indels 955; Gaps 134;
7 EEDGGAETQDSEDE-----EDEMEDDDSDYPMEEMEDDDDDASVCTESSPRSHS 59

Db	787	KNEKTDKERTFDERVERERRLTRKKEVKDKTDKO-----	822
Qy	60	TYSTPGRRKPRVRRSPILE-----EKDIPPLEFPKS-----SEDLMPVNEHI	104
Db	823	-----KRKGKVHSPSSQSSETDQENEREOSP-EKPRSCNKLRSREKADKGIKAKRLEL	874
Qy	105	MNVIAIYEVLRNFGTVLRLSPFRFDECAALVSQOCTLMAEMHVVLKAVLREEDTSNT	164
Db	875	MPCVWLTVRKEGKVIDHTPV-----EKLKAKLDNDTVKSSALDOKLOVSOT	922
Qy	165	TGFPADLK--DSVNSTLYFIDGM-TWPEVLRVYCESD---KEYHHVLPYQOEADYYPGPV	218
Db	923	EPAKSLSLKLESVMKVPKEKGLSSHVEV-----EKEGRUKARKHLKPSQAD-----GVS	974
Qy	219	ENKIKVLOFLVDQFLTNITARELMSE---GVTIQYDDHCRVCHKGLDLCCECTCSAVYHL	275
Db	975	AVDLEAKRRFAFADSNLKAQKQPEVKSSPEMED-ARVLSKKQPDVSSREVILLREG	1033
Qy	276	ECVKPPL--BEVPEDEWOCEVVAHKVPQVTDCAVIAEQNKPIRIHEPTGYDRSRKRYWF	333
Db	1034	EAPRKVPYRKILKRESKIKLDRLNTVASPKDC-QELAS-----ISVSGSRPSSD	1083
Qy	334	LNRL--LIBEDTENENEKKIWIYSYKVOLAELIDCLDRDYWEALCKLILEM---REI	388
Db	1084	LQARLGALESVENOVOSKKPIPKQPKQL-----QVLDOGGEREDV	1129
Qy	389	HR-HMDITEDLTNKARGSNKSFLLAANEEILSIRAKG-DIDNVKS---PETEKDKN	443
Db	1130	RKNYCSLRDETPERKSGOEKSHSVNTEKI-----GIDIDHTQSYRKQMSRRKQO	1181
Qy	444	TENDSKDAE-----KNREFEDQSL-----EKSDDKTPDDPEQKSEEPTEVG	488
Db	1182	MEMEIAKSEFGPKKDVEDYERRSLVHEVGPKPDVTDSPSPSKKRMHDVDFICTKR	1241
Qy	489	DGKNSYSANLGDNTNATSEET--SPS-----EGRSPVGC--LSETPDSSNMAEKKVA	537
Db	1242	ERNYRSRQLSED-----SERTGQSPVRHGSHFHEDEDPGSPRLLSVKGSPKVDK	1293
Qy	538	SELPQD---VPEBPNK--TCESSENTATTTSIQPNLENSNSSELNSESOSAKAA---	588
Db	1294	VLPSYNTIVRESLAFNPYDSRRQMDMAKILSVLNSEDELARNWDSQMKODAGRED	1352
Qy	589	-----DPENGERESHPTPYSIQEBEIVGDFTFSEKSTGE	620
Db	1353	VSPFNSIIKRDLSLRKSRVDRLEGPVSDSDEGEHKSHP-----RASA	1397
Qy	621	LSPEPGAGKACASTRIIPLRNPDSKLSQKSOQVAAAAHAANKLFKEG-KEVLVNSQ	679
Db	1398	LYE-----SSRLFLLRDRDKLRE-RDERLSSSL-ERNKFPYFALDKTIIPDTK	1445
Qy	680	GEISR---LSTKKEVIMKGINNYFKLGOEGKVRVYHNOY-----	716
Db	1446	ALLERAKSLSSRE-----ENWSFLDWSRFANRNKKDKKVDASAPRPPSWYMKKK	1498
Qy	717	-----STNSPALNKHQHRDHDKRHL-AHKFLCTPAGEFKWNGSVHGSKVLTI	770
Db	1499	KIRTDSEGMDDKDKHKEEQERQELFASRF-----LH-SSIFEQDSKRL-	1543
Qy	771	ITOLENNIPSSFF-----HPNWAHSHRANIKAVOMCSKPRFALALAILCAVKKPVVMLP	825
Db	1544	-QHLERKEEDSDPISGRIYKQOTSEGAN-----STTDSIQEPPVLF-	1583
Qy	826	WREFLGHTRLHRMTSIREEEKKVKKKEKQEBEETMOQATWVKYTFPVKHQVWKQKE	885
Db	1584	-----HSRFMELTRMQOKEREKQPKPEVEKQEDTENHP-----KTPSAPENKDS	1629
Qy	886	EYRV---TGYGGMWSWISKTHVYRFVPLKPLGNTNVNRYKSLGEG---TKNNNDENMDES	939
Db	1630	ELKTPPSVGPSTVVTLESAPSALDKTTGD-----KTVAPLVTTEKTVEPATVSEEA	1683
Qy	940	KCSRSPPKTKIEPDSSEKDEYKSDAAKA-----DQ-----	970
Db	1684	KPASEPAPAVEQLEQVLDLPPGADDPKEAMMPAGVEEGSGGDPDPPYLDKAPPTPGCASE	1743

Db 266 LQTDYATEKESADGLQ-----GETQLL-----VSRKSRPKSGSGHAVA 304
Qy 411 AAAN--EETILESTRAKKGIDIVKSPET-----EKDKNET-----ENDSKDAEKNR 455
Db 305 EPASPEQELDONKGRDVESQVTPSKAVGASFPPIYEPAKMKTVPQVTSQQONSQKHKNK 364
Qy 456 EEF-----EDOSLEKD-----SDDKTDDDPBQGSSEETPYGDKGNSVY--ANLGDNTN 504
Db 365 DLYTTGRRESVNLGKSEGFAGDKTLTPRKLSTRNRTPAKVEDAADSAATKPNLSSKTRG 424
Qy 505 A--TSEETSPB--GRSPVCLSET-----PDSSNMAEK-----KVASELPQDYPE 546
Db 425 SIPTDVEYLPTEIETHNEPFLTLWLTVQERKIQKDSLKPKGLGTAGOMCSGLPLSSV 484
Qy 547 BENKTCESNNTS-----ATTTSIQPNL--ENSNSSELNSSQSESAAK-----588
Db 485 DINNFGDSINSEGIPLKRRRVSGFGLRPELFDENLPNTPLRGEAPTARKSLVMHTP 544
Qy 589 -----DDPE--NGERES-----HTPVSIQEEIVG-----DFTSEKSTGELSESPG 626
Db 545 PVLKKIIEQPOPQSGKQSGSEIHVEVKAQSLVISPPAPSPRKTVPASDQRRRSCKTAPA 604
Qy 627 AGKGAS-----GSTRILITRLNPDS-----KLSQLKSQQVAAAHAANKLFREGKEVL 674
Db 605 SSSKSQTEVPKRGGERVATCLOKRVVISRSQSHDILQICSKRRSGASEAN-----LI 656
Qy 675 VVNSQGEISRLSTKK---EVMKG---NINNYFKLGQEGKYRV--YHNOYSTN-----S 720
Db 657 VAKSWADVVKLGAKQOTQVIKHGPQORSNMNRORPATPKPVGVEHVSQFSTGHANSFCT 716
Qy 721 FALNKHQIHREDHDKRR--HLAHKFCLTPAGEPK-----WNGSVHGSKVTLTSTLRL 769
Db 717 IIGRAHTEKVVHVPARPYRLNMFISNQMDFKEDLSGIAEMFKTPVKEQPQLT--STCHI 775
Qy 770 TITOLENNIPSSF-----FHPNASHRANWIKAVQCMCK--PREFALALAILCAVAK 819
Db 776 AINSNENLGGKFOQTDSGEPLLTSESFGNGVFFSAQNAKQPSD-----KCSAS 827
Qy 820 PYVMLPIWREFLGHT-----RLHRMTSIERE-----ERKEK 849
Db 828 P-----PLRQCIRENGNVAKTPRNTYKMTSLTETKTDTEPESKTVSTVNSRGRSTEFNR 883
Qy 850 VKK--KEKKQBEETM-----QOATWVKYTFPVKHQVHWKQGEYRVTYGCGSWI 898
Db 884 IQKLPVESKEETNTEIVECILKRGOKATLQO-----RRGEMKEI-----925
Qy 899 SKTHVYRVPKLPNGNTNVNYSLEGTKNMNDMESDK-----RKCS--RSPKKIKIE 951
Db 926 -----ERPFET--YKENIELKEN--DEKMKAMKRSTWQOKCAPMSDLDLKL 970
Qy 952 PDSE--KDEVKG-----SDAAKAGADQNMEDISKITEKKOODVKELLDSDDKPKCE-EP 1002
Db 971 PDELMKDTARQONLLQTDHAKAPKSEKGIKTKM-----PCQSLOP 1012
Qy 1003 MEVDDDMKTESHVNCQESSQVDVNVSE-----GHLRTS-----Y 1038
Db 1013 ----EPINTPTTKOOLKASLKGVGKVELLAVGKFTTSGETTHTHREPAGDGKSIRTF 1068
Qy 1039 KKKTK-----SKLDGLLERRIKQFTLEKORLEKIKLEGGIKGIGKTSNKSNLSESP 1093
Db 1069 KESPKOILDPAARVGM--KKWPRTPKEAQASLEDL---AGFKELFQTPGSPSESMTEK 1123
Qy 1094 VITKAGEQCSMRQEQSPNANDQPE--DLIQCSQSDSSVLRMSDPSTHTNKLYPKDR 1152
Db 1124 T---TKIACKSPPPESVDPTSTKQWPKRSLRKADVEEFLALRLKLTSPSAGKAMLTPKPA 1180
Qy 1153 VLDDVSI-----RSPETKCPQNSIEN-----DIEKVS DL 1183
Db 1181 GGDEKDIKAFMCTPVOKLDLAGTLPKSKROLQTPKEAQALEDLAGFKELFOTPGHTEEL 1240
Qy 1184 ASRGO-----EPTKSKTKGNDFIDDSKLASADDIGTLT--CKNKPLPIQE 1227
Db 1241 VAAGTKTKIPCDSQPSQSDPVDVTSTKQR-----PKRSIRKADVEGELLACRNLMP----1290

Qy 1228 ESDTIVSSSKSALHSSVPKSTINDRD-----ATPLSRAMDFPGKLGCDSESNSTLENSSDT 1282
Db 1291 -----SAGKAMHTPKPSVGEENDIILFVGTPVOKL-----DLTENLTGS 1329
Qy 1383 VSIQDSSEEDMIVONSNESISEQFRTREODVEVLEPLKCELVSGESTG--NCEDRLPVPKGT 1341
Db 1330 KRRPQTPKEEAQALLEDLTGFKELFQTPGHTEEA-----VAAKTKTMCPESPSPESAD 1382
Qy 1342 EANGKKPSQOKKLEERPVNKCSDOIKLKNTTDDKKNNENRESEKKGQRTSTPQINGKONKP 1401
Db 1383 TPTSTRQPKTPLEKRDVOKELSALK-----KLTQTSGETHTDVKVGGEDKS 1430
Qy 1402 KIYLGECLEKIEISRVVSGNV--EPKVNINIKIIPENDIKSLTVKESAIRFINGDVIME 1460
Db 1431 INAFRETAQOKLDPAASVTGSKRHPKTK--EKAQPLEDLAGW--KELFQTP-----VCT 1480
Qy 1461 DFNERNSETSHLLSSSDAEGNYRDSLETLPSTKESDSTQTTTPSASCSPESNVQVED 1520
Db 1481 DKPTTHEKTKIACRSQPD-----PVDTPSSKPSQSKSLRKVD 1519
Qy 1521 MEIETSEVKKYTSS-----PITSEESNLNDFIDENGLPINK---NENYNGESKR 1568
Db 1520 VEEFFALRKRTPPSAGKAMHTPKPAVSGEKIYA--FM--GTPVQKLDLTENLTGSKRR 1574
Qy 1569 KTVITEVTMTSTVA-----TESTKTVIKVEGDKQOTVVVSTENCACAKSTVVTIT 1616
Db 1575 LQTPKEAQALEDLAGFKELFQTRGHTESMT--NDKTAKVACKSQSDPDLDKNPASSKR 1631
Qy 1617 TTVKLTSPSTGGSDIISVKEQSKTVVTVTDSLTGTTGTLVTSMTVKEYSTRDKVK 1676
Db 1632 RLKTSLG--KYGVKEELLAVGLKTQTSGETHTHTTEPTGDKGSKMAFMESPKQILDSAS 1689
Qy 1677 LMKFSRPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLNDDOL-----1715
Db 1690 LTGSKROLRTPKGSSEVPEDLAGFTELFQTSHTKESMTNEKTTKVSYRASOPDLVDNPT 1749
Qy 1716 -----KKLARKGGIREVPYFNNAKPAIDLWPPSPRPTFGITWRYLQIVKSLAGVSL 1769
Db 1750 SSKPQKRSRLKADTEE-----EFLAFRKQTPSAG---KAMHTPKPAVGEK 1793
Qy 1770 MLR--LLWASRLWDDMAAKVPPGGSTRTTETSETETITTEILIKRRDVGVPYGR--1820
Db 1794 DINTELGTPVOKLDQPGNLP--GSNRRLQTRKEKAQALELT-----GFRELFQTPCT 1844
Qy 1821 -----FEYCIRKIIC--PIGVP--EPPKETPTPQRKGLRSS-----ALRKP RP-----1859
Db 1845 DNPTTDEKTKILCKSPQSDPADPTNTKQRPKRSLLKADVEEFLAFRLKLTSPSAGKAM 1904
Qy 1860 ETPKQOTGPVITETWVAEELELEWEIRAF-----AERVE-----KEKAQAVE 1900
Db 1905 HTPKAA-----VGEERK-----DINTFVGTPEKLDLLGNLPGSKRRPQTPKEKAKALE 1952
Qy 1901 QOAKKRLEQKQPTVIATSTTSSTTSTISPAQKVMVAPISGVSVTGTTKMWLTTKVGP 1960
Db 1953 DLAGEFKELFQTPGHTHEESMTDDKITEVSKSPQDPVPKTP-----TSSKQRLKLSLCK- 2005
Qy 1961 ATVTFQONKNFHQTATWVKQGSNGVVVQOKVIL--GIIPSSGTGTSQOFTTSFQPTA 2018
Db 2006 -----VGVEEVLVPVKLTQTSKTTQT-----2028
Qy 2019 TVTIRPNTSGSGGT-----TSNSQVI-----TGPIRPGMTVIRTP-----LQOSTLG 2061
Db 2029 -----HRETAGDGKSTKAFKESAKOMLDPANGTGMERP-----RTPKEEAQSLDLAGF 2079
Qy 2062 KAIIRTPVMVOPGAQVMTQIIRGQVPSTAVSAPNTVSSTPGQKSLTSATSTSIQSSA 2121
Db 2080 KELFOTPDHTESTTDDKTKI-----ACKSPPPESMDPTST-----2117
Qy 2122 SOPPPRQOQVKLT--MAQLTQLTQGHGNGOGLTVVIOGQGTQTLQOLIP-----QGVTV 2175
Db 2118 RRRPKTPLGKRDIIVEELSALKQLTQ-----THTTDKVPGEDEKGINV 2159

1069 KESPKQILDPAARVTGM--KKWRTPKPEEAQSLD---AGFKELFQTPGPEESMTDEK 1123
QY VITRAKECQSDMSRQEQSPNANDQPE-DLIOGCSQSDSVLRWSDPSHTNKLXPDR 1152
Db ---TKIACKSPPPESVDPTSTQWPKRSRKADVEEFIALRKLTPSAGKAMLTPRPA 1180
QY VLDDVSI-----RSPETKCPKQNSLEN-----DIEEKVSDL 1183
Db GGDEKIDKAFMGTPVKQLDLAGTLPGSKRQLOTPKEKAQALDLAGFKELFQTPGHTTEL 1240
QY ASRGO-----BPTSKYKGNDFIDDDKLASADDIGHLI-CKNKKPLIQE 1227
Db VAAGKTTKIPCDSPQSDPVDPTSTKQ-----PKRSIRKADVEGELLACRLMP----- 1290
QY ESDTIVSSSKSALHSSVPKSTNDRO-----ATPLSRAMDFFGKLGCDSESNSTLNSSDT 1282
Db SAGAMHTPKPSVGEEDKIIIFVGTVPQKL-----DLJENLTGS 1329
QY VSIQDSEEDMIVONSNEISIEQFTRREQDYEVLEPLKCELVSGESTG-NCEDRLPVKGT 1341
Db KRRPQTPKEEAQALDLAGFKELFQTPGHTTEA-----VAAGKTTKMPCESSPPESAD 1382
QY EANGKKPSQOKLEBRPVNKSDQIKLXNTDKKNENRESEKKGORTSTFIQINGKONKP 1401
Db TPTSTRQPKTPLEKRDYOKELSALK-----KLTQSGETHHTDKVPGGEDKS 1430
QY KIYLGECLEKTESRVSNGV-BPKVNNINKIIPENDIKSLTVKESAIRPFINGDVIME 1460
Db INAFRETAQKLDPAASVTGSKRHPKTK--EKAQPLEDLAGW--KELFQTP-----VCT 1480
QY DFNERNSETKSHLLSSSDAEGNYRDSLETLPSTKESDSTQTPPSACSPESNVQVED 1520
Db DKPTTHXHTTKIACRSQPD-----PVDTPSSKPSQSKRSRKVD 1519
QY MEIETSEKVKTISS-----PITSEESNLNDFIDENGLPINK--NENVNGESKR 1568
Db VEEEFALRKRTPSAGKAMHTPKPAVSGEKNIYA--FM---GTPVKQLDLNLTENLTGSKRR 1574
QY KTVITEVMTSTVA-----TESKTVIKVEKGDQTVVSVSTENGAKSTVTMTT 1616
Db LOTPKKAQALDLAGFKELFQTRGHTTEESMT---NDKTAKVACKSSQPDLDKNPASSKR 1631
QY TTVTKLSTPSTGSGVDIISVKEQSKTVVTTVTDLSLTGTLVTSMVTSKEYSTRDKVK 1676
Db RLKTSIG--KVGKBEALLAVGLTQTSGETTHTEPTGDKGSKMAFESPQKILDSAA 1689
QY LMKFSRPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLPNDDL----- 1715
Db LTGSRQLRTPKGSVEVPEDLAGFTELQTPSHTKESMTNEKTKVSYRASQPDLDVDTPT 1749
QY KKLARKGGIREVPYFNYNAKPALDIWPSPRPRTFGITWRYRLQTVKSLAGVSL 1769
Db SSKPQKRSRKADTEE-----EFLAFRKQTPSAG-----KAMHTPKPAVGEK 1793
QY MLR-LLMASLRWDDMAKVPPGGSTRTSETSTETITTEITLKRVDVGYGIR----- 1820
Db DINTELGTPVKQLDQGNLP--GSNRLQTRKEKAQALELT-----GPRELFQTPCT 1844
QY FEYCIRKIIIC--PIGVP-ETPKETPTPORKLRSS-----ALRKRKP----- 1859
Db DNPTDEKTKKILCKSQSDPADPTNTKQPKRSKLKADVEEFELFRKLTPSAGKAM 1904
QY ETPKQTPVILETVAEELELWEIRAF-----AERVE-----KEKAQAVE 1900
Db HTPKAA-----DINTFVGTVPVEKVLDDLNLPGSKRRRPQTPKEKAKALE 1952
QY QOAKRLEQOKPTVIATSTTSSTSTISPAQKVMPAPISGSVTTGKWLITTKVGPSP 1960
Db DLAGFKELFQTPGHTTEESMTDDKITEVCSKSPQDPVKTP-----TSSQKRLKISLKG- 2005
QY ATVTFOQKNFHQTFATWVKQGSNGVVOQKVL--GIIPSSGTGTSQOTSTPQPRTA 2018
Db -----VGKVEVLVPGKLTQTSKTTQT----- 2028

QY TVTIRPNTSGSGT-----TSNSQVI-----TGPOIRPGMTVIRTP-----LQOQTGLG 2061
Db ---HRETAGDGKSIKAFKESAKQMLDPANYGTGMRWP-----RTPKEEAQSLDLAGF 2079
QY KAIRTPVMQVOPGAQVMTQIIRGQPVSTAVSAPNTVSTPGOKSLTSATISNISSA 2121
Db KELFQTPDHTTEESTDDKTKI-----ACKSPPPESMDTPTST----- 2117
QY SOPRPOGOVKLH--NAOLTQLTQGHGNOGLTVWIOGOGQTTGQLQLIP-----QGVTV 2175
Db RRRKPTPLGKRDIVEELSALKQLHQ-----TTHTDKVPGEDDKGINV 2159
QY LPGPGQQLM-----QAAMPNGTVQ-----RFLF-TPLATATTATTTTTTV 2215
Db FRETAKQKLDPAASVTGSKRQPRTPKGAQPLEDLAGLKFELFQTPCTDKPTTHEKTTKI 2219
QY STTA-----AGTGE--QROSKLS--PQOVHQDKTLPPAQSSSVGPAKAQOTA----- 2260
Db ACRSPQDPVPVGTPTIFPQSKRSRLKADVEESIALRKRTPSVGKAMDTFAPAGGDEKDM 2279
QY QPSARPOPOQOPQSPAQPEVQTPQEVQTTQTVSS-----HVPSEAQPT 2303
Db KAFMGTPVKQLDLPGNLPGSKRWQTPKE-----KAQALDLAGFKELFQTPGDKPT 2332
QY HAQSSKPQAAQSQPOQSNVQSPVQSPQTRIRPS-----TPS- 2344
Db TDEKTTKIACKS-PQ-----PDVDTPASTRQRPKRNLKADVEEFIALRKRTPSA 2383
QY QLSPGQSQVQITTSQPIQPHTSI-QIP-SOGQPOQSP----- 2382
Db GKAMDTPKPAVSDENKINIVFETPVKQLDLGNLPGSKRQPOTPKEKAEALDVLGPKEL 2413
QY QVOSSTQTLSSGQTLNOVSVSPSRPQIQOQPQVIAVP-----QLOQOVQVLSQI-- 2435
Db FQTPGHTTEESMTDDKIIEVCSKSPQESFKTSRSKORLKLPLVKVDMKEEPLAVSKLTR 2503
QY QSVVVAQIAOQSQVPOQIKL--QLPQIQOSSAVQHOIONVTVQAAVQEOQLORVOQ 2493
Db TSGETTQTHTEPTGDSKSIKAFKESPKQILDPA-----SVTGSRRQLTRKEKARA 2555
QY LRDOQOKKK-----QOOIEINVN-----TPSKLLIKVEIIOKQVVM 2529
Db LEDVDFKELFSAGHTEESMTIDKNTKIPCKSPPELTDATSTKRCPKTEL--RKEVK 2613
QY KHNAVIEHLKQKKSMT-----PAERENQRMIVCNQVMKYLDKIDKEEKQAAKKRKE 2583
Db EELSAVERLTQTSQSTHTEKPEASGDEGIKVL-----KQRAKKKPNP 2656
QY ESVEOKRSKQNA-----TKLSALLFKHKEOLRA-----EILKK 2616
Db VEEPSRRRRPRAPKEKAQPLEDLAGFTELSETSGHTQESLITAGKATKIPCESPPLVVD 2716
QY RALLDKQIQIEVQELKRDALKIKKEDLMQLAQ-----AT 2651
Db TASTKRHLRTRVQK-----VQVKEPSAVKQTSGETTDDADKEPAGEDKIGALKESAK 2771
QY AVAAPCPVTPVLVAPPAPPPSP-----PPGVQHOHTGLLSTPTLPVASQKRRKEE 2703
Db QTPAPAAVTSRRRPRAPRESAQAIEDLAGFKDPAAGHTEESMTDDKTTKIPCKSSPEL 2831
QY KDSSSKSKKK-----KMISTTSKETKDKTKLYCKICKTQPVDESK 2741
Db EDTATSSKRRRPRTRAQKVEVKEELLAVGLKLTQTSGETTHTD-----KEPVGEK 2880

RESULT 7
US-09-864-761-34248
; Sequence 34248, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.

```

1  APPLICANT: Hanzel, David K.
2  APPLICANT: Chen, Wensheng
3  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
4  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
5  FILE REFERENCE: Aecomica-X-1
6  CURRENT APPLICATION NUMBER: US/09/864,761
7  CURRENT FILING DATE: 2001-05-23
8  PRIOR APPLICATION NUMBER: US 60/180,312
9  PRIOR FILING DATE: 2000-02-04
10 PRIOR APPLICATION NUMBER: US 60/207,456
11 PRIOR FILING DATE: 2000-05-26
12 PRIOR APPLICATION NUMBER: US 09/632,366
13 PRIOR FILING DATE: 2000-08-03
14 PRIOR APPLICATION NUMBER: GB 24263.6
15 PRIOR FILING DATE: 2000-10-04
16 PRIOR APPLICATION NUMBER: US 60/236,359
17 PRIOR FILING DATE: 2000-09-27
18 PRIOR APPLICATION NUMBER: PCT/US01/00666
19 PRIOR FILING DATE: 2001-01-30
20 PRIOR APPLICATION NUMBER: PCT/US01/00667
21 PRIOR FILING DATE: 2001-01-30
22 PRIOR APPLICATION NUMBER: PCT/US01/00664
23 PRIOR FILING DATE: 2001-01-30
24 PRIOR APPLICATION NUMBER: PCT/US01/00669
25 PRIOR FILING DATE: 2001-01-30
26 PRIOR APPLICATION NUMBER: PCT/US01/00665
27 PRIOR FILING DATE: 2001-01-30
28 PRIOR APPLICATION NUMBER: PCT/US01/00668
29 PRIOR FILING DATE: 2001-01-30
30 PRIOR APPLICATION NUMBER: PCT/US01/00663
31 PRIOR FILING DATE: 2001-01-30
32 PRIOR APPLICATION NUMBER: PCT/US01/00662
33 PRIOR FILING DATE: 2001-01-30
34 PRIOR APPLICATION NUMBER: PCT/US01/00661
35 PRIOR FILING DATE: 2001-01-30
36 PRIOR APPLICATION NUMBER: PCT/US01/00670
37 PRIOR FILING DATE: 2001-01-30
38 PRIOR APPLICATION NUMBER: US 60/234,687
39 PRIOR FILING DATE: 2000-09-21
40 PRIOR APPLICATION NUMBER: US 09/608,408
41 PRIOR FILING DATE: 2000-06-30
42 PRIOR APPLICATION NUMBER: US 09/774,203
43 PRIOR FILING DATE: 2001-01-29
44 NUMBER OF SEQ ID NOS: 49117
45 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
46 SEQ ID NO 34248
47 LENGTH: 2665
48 TYPE: PRT
49 ORGANISM: Homo sapiens
50 FEATURE:
51 OTHER INFORMATION: MAP TO AL034555.2
52 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
53 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
54 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
55 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
56 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
57 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
58 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
59 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
60 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
61 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
62 OTHER INFORMATION: EST_HUMAN HIT: AU117052.1, EVALUE 0.00e+00
63 OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUE 3.00e-10
64 US-09-864-761-34248

```

[illegible]

Query Match	2.7%	Score 402;	DB 10;	Length 4019;
Best Local Similarity	17.34;	Pred. No. 1.8e-11;		
Matches 510;	Conservative 314;	Mismatches 992;	Indels 1126; Gaps 108;	
QY	254	CRVCHKL---GDLCCETCSAVYHLECVKPLEEVPEDEMOCEVCAHVKGVTDCVAEI	310	
Db	65	CEAGKATDGRLLLCDDCDSIYHYCLDPLQTPVGGWKCKWCWCRHCGATSAGLRC	124	
QY	311	QKNRPYIRHEPIGYDRSRKRYFLNRRLLIIEEDTENENKKIWIYSTKVQLAELIDCLDK	370	
Db	125	EWNNYTQCAPCA-SLSSCPVCYRNYR---EEDL-----ILQCRQC	161	
QY	371	DYWEAECLKILEEMREELHRMDITEDLTNKARSGNSKFLAANEELIESIRAKKGIDIN	430	
Db	162	DRMHAVQCNL-NTEEEVENYADIGFDCS-----MCRPYMPAS-----N	199	
QY	431	VKSPEETKDKNETENDSKOAEKNREEFEDQSLKSDDKTPDDDPQKSEETEVGDK	490	
Db	200	VPSDCCB-----SSSLVAQIVTKVELDPPKTYTQDGV-----	232	
QY	491	GNYSANLGDNTTATSEETSPSEGRSPVGCSETPDSSNMAE-KKVASELPQDVPPEPN	549	
Db	233	-----CLTE-----SGMTQLQSLITVTVPRKRSKPK	258	
QY	550	KTCSSN-TSATTTSIQPNLNSNS-SSELNSSQSESAAADDPENGERSHTPTVSIOEE	607	
Db	259	LKLKIINONSVAVLQTPDIOSEHSRDMGDDSRGELMDCD-----GKSES-----SPERE	310	
QY	608	IVGFTSEKSTGELSEPGAGKAGSGSTRIITRLRNPDLSKLSQKSOQVAAAHEANKLF	667	
Db	311	AVDDET-----KGVEGTDGVKKRRKP-----Y	333	
QY	668	KEGKVLVYNSQGEISRLSTKKEVIMK---GNINNYFKLGOEGKYRYVHYNQYSTNSFALN	724	
Db	334	RPGIGGVMYRQRRTGQGTKRSVIRKDDSSGSISEQLPCRDDG-----	376	
QY	725	KQHREDHDKRRHLAHLFCLTPAGEFKWNGSVHGSKVLITSTLRLTITQLENNIPSPFFH	784	
Db	377	-----	376	
QY	785	PNWASHRANWIKAVQCMCKPREFALALALECAVKPVVMLPIWREELGHTLRHMTSIER	844	
Db	377	-----WSEQLPDTLVDGESSV-T	393	
QY	845	EKEKVKKKKKQBE--EET---MQATWYKYTFPVKHQVWKQKGEEYRTVGGWSWI	898	
Db	394	ESTEKIKRKRKNKLEETFPAYLQEAFFCKDLDDTSRQ-----SKI	436	
QY	899	SKTHYRVFVKLPQNTNVNRYKSLGKTKNNMDENDESDDRKCRSPKKIKIEPDSEKDE	958	
Db	437	SLDNLSEDAQLLYKTNNN-----TGFLDPSLD--PLLSSSSAPT-K	475	
QY	959	VKGSAAKAGDQNEWDISKITEKKDQVKELLDSDSKPKCEEPW-EVDDDMKTESHVNC	1017	
Db	476	---SethGAPDPLADISEVL-NTDDDLGLIISDLAKSVHDIGPVTDPPSLPQPNV	531	
QY	1018	QESSQDVVNVSEGFHLRTSYKKTKSKLDGLLERIKQFTLEEKORLEKIKLEGGIGK	1077	
Db	532	NOSS-----RPLSEQLDGL-----SPELDKMTVDGAILG	562	
QY	1078	-IGTSTNSSKNLSSESPVITAKEGCQSDSMRQEQSP-----NANNQDEPDIQGCQ	1129	
Db	563	KLYKIPLEIGGRDVEDLEFATVLSPANTQPTPLPQPPPTQLPIHNQDAFMRPLMNLIG	622	
QY	1130	SDSSVLRMSDP-----SHTTNKLYPKDRVLDVSIIRSPETKCPKQN-----SIEND	1175	
Db	623	SSPHLPNSLPPEGGLGTFESAIAOSSYPDAR--DKNSAFNPMASDPNNSWTSSAPTVEGE	680	
QY	1176	IEEKVSLASRGQEPKTSKTKGNDFFIDDSKLASADDIGTLCKN-KKPLIOESDITVS	1234	
Db	681	-----NDTMSNAQSTLKWEK-----EALGEMATVAPVLYTNINFPNLKEEFPDWT	728	
QY	1235	SKSALHSSVPKSTNDRDATPLSRAMDFEGKLGCD--SESNSTLSENSDVTSTQDSSSE--	1290	

Db 1646 GNFIPRPDPGPRHTDMRRPQGLPNQLPVHPDLEQVPPSQEQGHSHSSMMVMTLN 1705
Qy 2100 -----SSTPGOKSLTSATSTSNK-----SOPRRPOGOVK----- 2133
Db 1706 HPLGGESEAPLSTVSPSTSDNLOITTPQSDGLEKLDSDPSTVKELDVADLEGVEVK 1765
Qy 2134 -----LTAQTLQTLQTHGGNOGLTVVIGOGQTTGQLQLIP----- 2170
Db 1766 DLDDDELENLNDTEGKVELDTLONLETNPNLDDLLR-----SGEFDIATYDPELD 1820
Qy 2171 -----OGVTVLPGPGQQLMQAAMPNGTVORFLFTPLATATTA 2208
Db 1821 MGDKGMFNEELDLPIDDLNOCVSEPKKEQ-----ENKTLVLDKHSPOKK 1870
Qy 2209 SITTTIVSTAAAGTGOROSKL-----SPQOVHQDKTLTPAOSSSVGPA-- 2253
Db 1871 STVINEKTEVLSPNSKVESKETENKDNKNVDTPCQASAHSDNDGKTSLHPCDP 1930
Qy 2254 -----KAQPTAQPSPARPOTOPQSPAPEVQTOPEVOTOTVSSHVPSEAOPT---HA 2305
Db 1931 DLPEKRTNRETAGPSAN-VIASTQLPAQDVINSCGITGSTPVLSSLLANEXSDNDIRP 1989
Qy 2306 QSKPQVAQSQPOSVQOSPVQSPQTRIRPSTPSQLSPGQ-----QSQ 2353
Db 1990 SGSPPTTLPASPSNHVSLPPF-----IAP-----PGRVLDNAMNSVTVVSR 2033
Qy 2354 VQTTTQPIPIOPHTSLQIPSO-----GPOSO---PQV-QSSTQTLSSGOTL---N 2398
Db 2034 VNVFSGQGVNPLG---IPGQSTVNHSLGTGKPATGTGPTQSQTSSMSGPOQLMIPQ 2090
Qy 2399 QVSVSSPSPOLQIQOPQPVIAVPOLOQOVVLSIQSOV----- 2439
Db 2091 TLAQONRERPLLEEOPLLLOLDLQEROEQOQOQOMQAMIRSEPPFPNIDFADITDP 2150
Qy 2440 -----VAQIOAQS-GVPOQIKLQLPQIOQSSAVQ----- 2469
Db 2151 IMKAKMVALKINKVMAQNLGMPVMVSRFPMPGVGTGTONSEGNLGPQAIPDGSI 2210
Qy 2470 THQION-----VVTQAAASVOEQLRVQOQRDQOQKKQOQIENVNTPPSKLLIK 2519
Db 2211 THQISRPNPNFPGFVNDORSQYEEWLEQTOQLLOMQOQYLEEQIG----- 2258
Qy 2520 VEIQOVKMNHNAVIEHLKQKSMTPAEREENQRMVNCVMKYILDKIDKEEQAANK 2579
Db 2259 -----AHRKSKKALSAKOR-----TAKK 2276
Qy 2580 RKREESVEQRKQONATKLSALLFKHKEQLRAELKRALDKLDLOIEVQEELKRDLIK 2639
Db 2277 AGREPPEDEQLKHVTEQOSMVQOLEQIRK-----OQKEHAELIEDYRIK 2323
Qy 2640 KERDLMQAAATAAAP-CPVPTVPLPAPPAPPPPPPGV-----QHTGLLSTPTL 2691
Db 2324 QQQ---QCAMPTMPSVQPPPLIPGATPPTMSQPTFPVQLOHQOHTTVISGHTS 2380
Qy 2692 PV 2693
Db 2381 PV 2382

RESULT 9

US-09-854-133-425
; Sequence 425, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133

Query Match 2.7%; Score 402; DB 10; Length 4019;
Best Local Similarity 17.3%; Pred. No. 1.8e-11;
Matches 510; Conservative 314; Mismatches 992; Indels 1126; Gaps 108;
Qy 254 CRVCHKL---GDLLOCETCSAYVHLECYKPPLEEYDEDEWQCEVCAHVKPVGTVDCVABI 310
Db 65 CEACGKATDPGRLLLCDDCDISYHTYCLDPPLOTVPKGGWKCKWCVCRHCGATSAGLRK 124
Qy 311 QNKPKYIRHEPTGYDRSRKRYWFLNRLLIIBEDTENENKKTWYISTKVQLAELIDCLDK 370
Db 125 EWQNNYTQCAPCA-SLSCPCVCYRNYR---EEDL-----LLOQKOC 161
Qy 371 DYWEAEELKILEEMREEIHRHMDITEDLTNKGKSNKGFLLAAANEELLESIRAKKGDIDN 430
Db 162 DRMHAVCONL-NTEEEVENVADIGFDCS-----MCRPYMPAS-----N 199
Qy 431 VKSPETEKKDNETENDSKDAEKNEEFEDQSLKSDDDKTPDDDDPEOGKSEEPTEVGDK 490
Db 200 VPSDDCE-----SSSLVAQIVTKVKELDPKPKTYTODGV----- 232
Qy 491 GNSVSNALGDNTNATSEETSPSEGRSPVGCLESETPDSSNMAE-KKVASSELQDVPPEPN 549
Db 233 -----CUTE-----SGMTQLQSLVTVVPRKRKPK 258
Qy 550 KTCESN-TSATTTSIQPNLENSN-SSELNSSQSEAKAADPENGERSHPTVSIQEE 607
Db 259 LKLIINQNSVAVLQTPPDIOSEHSRDGEMDSDREGELMDCD-----GKSES-----SPERE 310
Qy 608 IVDFTSEKSELSESPGAGKAGSGSTRITIRLNPDSKLSQLSQQAHAHANKLF 667
Db 311 AVDDET-----KGVEGTGDKRKRKP-----Y 333
Qy 668 KEKGEVLVNSOGEISRLSTKKEVIMK---GNINNVFKLGQGVKRYVHNQYSTNSFALN 724
Db 334 RFIGGFMVQRSGRTQGGTKRSVIRKSSGSIQELPCRDDG----- 376
Qy 725 KHQREDHDKRRHLAHKFLCTPAGEFKWNGSVHGSKVLITSLRLTITOLENNIPSSFFH 784
Db 377 ----- 376
Qy 785 PNWASHRANWIKAVQCMKXPREFALALILECAVKPVVMLPIWREFLGHTRHMTSIER 844
Db 377 -----WSEQLPDTLVDESVSV-T 393
Qy 845 EEKEKVKKKEKOE--EET---MQOATWVKYTFPVKHQVWKQKGEERYVTGYGGSWI 898
Db 394 ESTEKIKKRYKRKNKLEETFPAYLQEAFFGKDLDTSRQ-----SKI 436
Qy 899 SKTHVYRFVFKLPNTNVNRYKSLGKTKNNMDENMDSDKRCSPKKIKIEPDEKDE 958
Db 437 SLNLSGEDGALLYKTNMN-----TGFLDPSLD--PILLSSSSAPTK----- 475
Qy 959 VKGSDAAKADQONEMDISKITEKKDQDVKELSDSDKPKCKEPM-EVDDDMKTESHVNC 1017
Db 476 ---SGTHGPAADDPLADISEVL-NTDDILGIISDDLAKSVDSHSDIGPVTDTPSSLPQPNV 531
Qy 1018 QESSQDVVNVSEGFFHRTSYKKKTKSSKLDGLLERRIKOFTLEEKQRLKLEKIGK 1077
Db 532 NQSS-----RPLSEQLDGLL-----SPELDKMVTDCAILG 562
Qy 1078 -IGKTTNSKNLSESPVITKAKEGQOSMRQEQSP-----NANNQPEDLIQGCQ 1129
Db 563 KLYKIPELGGKDVEDLFTAVLSPANTQPTPLPQPPPTQLLPIHNDASFMRPLMGLIG 622

		Matches 221; Conservative 91; Mismatches 418; Indels 254; Gaps 33;	
QY	1780 WDMAAKVPPGGGSTRTET-SETETITTEIIKRRDVGPGYRFEVCIRKIIICPIGVP--E 1836		
Db	3674 ---TTTTVTPPTPTGTGTPTTTTPTTTTTPPT-PTGTQ-----TPTTTPTT 3720		
QY	1837 TPKEPTPORKGLRSSALRP-----KRPETPKQTG-----PVIIETWAEELWEI 1884		
Db	3721 TTTVTPPTPTGTGTPTTTTPTTTTTPPTPTGTGTPTTPTTTTPTT----- 3771		
QY	1885 RAFAEERVEKEKAQAEVQOAKRLEQKPTVIATSTPTS---STSTISIP-----AQK 1935		
Db	3772 -----PTPTPTGTGTPTTTTPTTTTTPPTTPTGTGTPTTPTTPTGTGT 3804		
QY	1936 VMVAPISGSVT-----TGKMWLTTKVGPATVTFQONKHFQTFATWVKOGOSNGV 1988		
Db	3805 PTTTPTTTTTPPTPTGTGTPTTTTPTTTTTPPTPTGTGTPTTPTTPTTPTT 3861		
QY	1989 VQVQOKVLGIIPSSGTSTQOFTSFQPRATVIRPNTSGSGGTTSSNOVITGPIRPGM 2048		
Db	3862 -----VTPTPTGTGTPTT-TPITTTTTPPTPTGTGTPTTPTTPTTPTT 3908		
QY	2049 TVIRTPLOQSTLGKAIIRTPVMVQGAQVQVMTQIIRGQPVSTAVSAPNTVSSTPGQKSL 2108		
Db	3909 TPPTPTGTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 3963		
QY	2109 TSATSTSNIOSSASOP-PRPQOGQVKLTMAQLTQLTQGHGGNGGLTVIQQOGQTTGQLQ 2167		
Db	3964 QPTPTTPTTTTTPPTPTGTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 4015		
QY	2168 LIPOGVTVLPGGQQLMQAMPNGVQRFLEPLATT-----ATTASITT 2212		
Db	4016 PTTTTTTPPT-----TPTGT-OTPTTPTTTTTPPTTPTTPTTPTTPTTPTT 4066		
QY	2213 TTVSTTAAGTGEORQSKLSFQMVHODKTLPPAQSSSVGCPAKAQOTAPSAARPQPTQP 2272		
Db	4067 TTVTPPTPTGTGTPTT-----TPTTPTTTTTPPTPTGTGTPTTPTTPTTPTT 4114		
QY	2273 QSPAPEVOTQPEVOTQTTVSSHVPSEAQPTHAQSSKPKQVAAQSQPSNVQCSQSPRVQS 2332		
Db	4115 VTPPTPTGTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 4172		
QY	2333 PSQTRIRPS---TPSQLSPGQSQVOTTTSQPIPTQPHLSLOIPSGOQPO-----SOPQV 2384		
Db	4173 PTTTPTTTTTPPTPTGTGTGPPHTSTAPIAELTTSNPPPSSTPTQTSRSTSSPLT 4232		
QY	2385 QSSQTQLSSQTLNQVSVSSPSRP 2408		
Db	4233 ESTLLSTLPPAIENTSTAPPSTP 4256		

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:12:14 ; Search time 23012 Seconds
(without alignments)
17537.523 Million cell updates/sec

Title: US-09-698-295-2
Perfect score: 9865
Sequence: 1 ggccaggctgagtgccgc.....acaaaaaaaaaaaaaaaa 9865

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_higo_hum.*
- 40: em_higo_mus.*
- 41: em_higo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	9067	91.9	9700	9	AB032251	AB032251 Homo sapi
2	7836.4	79.4	8295	9	AY282495	AY282495 Homo sapi
3	3033	30.7	3419	9	AK025776	AK025776 Homo sapi
4	2565.6	26.0	2673	9	HSU05237	U05237 Human fetal
5	2316	23.5	88848	9	AC107377	AC107377 Homo sapi
6	2316	23.5	108860	9	AC127525	AC127525 Homo sapi
7	2314.4	23.5	199875	9	AC134407	AC134407 Homo sapi
8	1998	20.3	4074	9	HS080468	BD157797 Homo sapi
9	1719.4	17.4	2057	6	BD157797	BD157797 Primer fo
10	1719.4	17.4	2057	9	AK022570	AK022570 Homo sapi
11	1714	17.4	2549	10	BC032289	BC032289 Mus muscu
12	1708.4	17.3	2769	10	BC046973	BC046973 Mus muscu
13	1690.2	17.1	2481	10	BC037661	BC037661 Mus muscu
14	1496.2	15.2	2214	10	BC021489	BC021489 Mus muscu
15	1454.4	14.7	23387	2	AC109873	AC109873 Rattus no
16	1454.4	14.7	238226	2	AC120467	AC120467 Rattus no
17	1454.4	14.7	280450	2	AC111698	AC111698 Rattus no
18	1448.4	14.7	217735	10	AL596116	AL596116 Mouse DNA
19	1394.6	14.1	1606	9	BC050566	BC050566 Homo sapi
20	1240.2	12.6	1540	6	AX552123	AX552123 Sequence
21	1190.6	12.1	1576	9	BC037422	BC037422 Homo sapi
22	1071.2	10.9	1955	10	BC003956	BC003956 Mus muscu
23	952.2	9.7	1976	9	AK027184	AK027184 Homo sapi
24	831	8.4	179759	2	AC132868	AC132868 Homo sapi
25	831	8.4	191280	2	AC087389	AC087389 Homo sapi
26	831	8.4	193519	2	AC023123	AC023123 Homo sapi
27	814.6	8.3	1032	6	AR216196	AR216196 Sequence
28	771.4	7.8	841	6	BD147323	BD147323 Primer fo
29	688.2	7.0	62138	2	AC100831	AC100831 Homo sapi
30	662.6	6.7	60526	2	AC130323	AC130323 Homo sapi
31	627.4	6.4	750	6	BD018361	BD018361 Novel gen
32	627.4	6.4	750	6	BD098299	BD098299 Novel gen
33	598.2	6.1	183093	9	AC138688	AC138688 Homo sapi
34	598.2	6.1	187685	9	AC127032	AC127032 Homo sapi
35	598.2	6.1	190285	2	AC126604	AC126604 Homo sapi
36	598.2	6.1	190844	2	AC073990	AC073990 Homo sapi
37	586.2	5.9	139585	9	AC090419	AC090419 Homo sapi
38	586.2	5.9	170030	9	AC005829	AC005829 Homo sapi
39	586.2	5.9	226696	9	AC139026	AC139026 Homo sapi
40	585.8	5.9	134016	2	AC048388	AC048388 Homo sapi
41	574	5.8	574	6	AX677732	AX677732 Sequence
42	574	5.8	2123	6	BD159315	BD159315 Primer fo
43	574	5.8	2123	9	AK022081	AK022081 Homo sapi
44	547.6	5.6	80206	2	AC138693	AC138693 Homo sapi
45	538.2	5.5	556	6	BD151960	BD151960 Primer fo

ALIGNMENTS

RESULT 1
AB032251
LOCUS
DEFINITION Homo sapiens BPTF mRNA for bromodomain PHD finger transcription factor, complete cds.
ACCESSION AB032251
VERSION AB032251.1 GI:6683491
KEYWORDS bromodomain PHD finger transcription factor; BPTF.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Jones,M.H., Hamana,N. and Shimane,M.
TITLE Identification and characterization of BPTF, a novel bromodomain

transcription factor
Genomics 63 (1), 35-39 (2000)
20130111
10662542
2 (bases 1 to 9700)
Jones,M.H., Hamana,N. and Shimane,M.
Direct Submission
Submitted (08-SEP-1999) Michael H Jones, Chugai Institute for
Molecular Medicine, Senior Research Scientist; 153-2, Nagai,
Niihari, Ibaraki 300-4101, Japan (E-mail:mike@immed.com,
Tel:81-29-830-6211(ex.271), Fax:81-29-830-6270)
Location/Qualifiers
1..9700
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q23-24"
/tissue_lib="testis"
1..9700
/gene="BPTF"
472..8817
/gene="BPTF"
/codon_start=1
/product="bromodomain PHD finger transcription factor"
/protein_id="BAAB9208.1"
/db_xref="GI:6683492"
/translation="MVSEEEEDGAEETQDSEDDDEDEDDDDSDYPEEEDDD
DDASYTESFRSTSYSTPGRRKPRVHRPILSEEDIDPLEPESKEDLMVNE
HIMVIAIEYLRNFGVLRISLRFDFCAALYQEOCTIMAEHMHVYLKAVLRVE
TSNTFGPADLQSVNLTLYFDGMTWPEVLYRVYCESDKYHHVLPYQEAEDYPGPV
ENKIKVLQFLVDQFLTNIAREELMSGVLYQDHCVRCHKLGLDLCETCSAIVHLE
CVKPLEVPEDEQCEVCAHKVPGYDCVAETQKNKPYIRHEPIGYDRSRKYWFL
NRLIIEEDTENENKKIWIYSTVKQLAELIDCLDLDYWEAEELKILFEEREIHRHM
DITDELTKNAKSNKSLAANBEILSIRAKKGDIDNVKSPETEKDKNETDNDSD
AEKREEFDQSLKSDDKTPDDPQGESEVDFKSEKSNGLSESPGAGKASS
TRITRLRNPDQSKLSQKSOQAAAAHEANKLFKEGEVLVYNSQGSLSRLTSKEVI
MKNINNYKLGQEGKYRVYHNOYSTNSFALNHQHRHEDHDKRHLAKFCLTPAGFE
KNGSVHGSKVLTLTSLRTLTOLENNIPSSFLHPNASHRANIKAVQMSKPREFA
LALALECAKPVVMLPIWREFLGHTRLRHTSITEREKEKVKKKKKQBEETMQQA
TWKTYFPKQVWKQGEERYVYGYGWSHSIKTHYRVFVPLPGNTNRYKSLGG
TKNNDENDDSKKRSRKKIKIEPDSKDEKVSQDSSQVQVNVNVSNGHLRSYIK
DDYVELLSDSDPKPEEPEVDDDMKTESVHNCQESSQVQVNVNVSNGHLRSYIK
KTKSKLDGLLERRIKFTLEEKQLEKIKLEGGIKGIGTSTNSKNLSPEVITKA
KEGQSDSMROEKNANNDQEDLIQCSQSDSSVLRMSDPSTTNKLYPKRVLDD
VSIKSPETCKPONSIEDIEKVSILASRGQETKSKTKGNDFFIDDKLASADDIG
TLICAKKPLQESDITVSSKSAHSSVPKSTNDRDAPLSRAMDFEGKIGCDSES
NSTLENSDITVSIQSSSEDMIVQNSNESISEQFTRDQVLEPLKCELVGSESTG
NCEDLRPVKTEANGKPKQKLEERPVNKCSDQIKLNTDKKNENRESEKQOR
TSFQINGDKNPKIYLGECLEISESVVSGNVEPKVNNINKIIPENDIKSLTYRE
SAIRPFINGDVLMEFNRSSETKSHLLSSDAEGNYRDSLETLPKESDSTQTTT
PSACPESNVNOVDMIEIETSEVKVTSSPITSEESNLNDFIDENGLPINKENY
NGESKRTVIEVTMTSTVATESKTVIKVEKGDKQTVVSTENCAKSTVTMTT
KLSSTSGSDVLIISVKQSKVTVTVTDSITTTGGTLVTSMTVSKYSTKRVKLM
KFSRPKKTSRGTALPSYRKRYKTKSKSIFVLVPLNDMLKKLARKGIREVPYENAKP
ALDIMPSPRPTFGITRWYRLQTVKSLAGVLSLWLRLLWASLRMDMAAKVPPGGST
RFTSETTEITTEILKRRDVGPIGYEFYKIKIICPIGPETPKETPORKGLRSS
ALRKRPKPTKQTPVITETWVAEEELWEIRAFPAERVEKEKAQAPQAKRLEQQ
KPTVATSTPSSTSTISPAQVMVAPISGSVTTCTKMTLTKVGSPTATVTFQON
KNFQTPATWVKQSGSVQVQKVLGILIPSTGTSTQOTFTSPQRTATVTRPNT
SGSGGTSNSQVIGTQIRPMGVITPLQOSTLGAIIIRPVVMVQAPQVMTOL
ROGPVSTAVSAPNTVSSTPGKSLTSATSNISQSSASQPPRQPGQVKLTMAOLTOL
TQGHGNGQLTVIIOGGQGTQGLQILIPGVTVLPQGGQQLMQAMPNGTVQREFLTP
LATATASITTTTTTSTVTAAGTEQKSLSPQMVHQDKTLPPAASVSSGPAKAAQO
TAQPSARPOPTQPSAPQAEVQTOPEVQTOFTVSHVSEAPTHAQSSQVAAQS
QPSNVQGSVPVRQSPQSRIRPSTPSQSPQSOVQTOFTTSPITQPHQSLQIPS
QGPQSQPVQSTQTLSSQOTLNQSVSSPSRPOLOIQOPQVIAVPOLOQOVVL
SQIQSQVQVIAQQSGVFPQIKLQPLQIQOSSAVQTHQIQNVVTVQAAVSQVQELQ
VQQLRDQQKQKQKQIQIKREHTLQASNOSEIIQKQVNMKNHVAIEHLKQKSWTPAE
REBNQRMVNCVMYKILDKIDKEKQAKRKEESVEKRSKQNAKTLKSALLEFKHK
EQLRAELKKRLDKLQIEVQELKRLKIKKEDLQMLAQATAVAPCPVPTVL
PAPPAPPPSPPPGVQHTGLLSTPLTPVASOKRREKEDSSSKSKKMKMISTTSKE
TKDTKLYICKTTPYDESKFYICGDRQCNWYHRCVGLIQSEAEILIDEVVCQCCQSTE

polya_site 3235 a 2127 c 2199 g 2139 t
BASE COUNT
ORIGIN
Query Match 91.9%; Score 9067; DB 9; Length 9700;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 9480; Conservative 0; Mismatches 5; Indels 380; Gaps 3;
DAMTVLTPLTEKDYEGLRKRVLRSLQAHKMWPFLEPVDNDPDIYGVIKPEMDLATM
EERVQRRYIEKLTETFEVADMTKIFDNCRYNPSDFYOCAEVLESFFVQKLGKFASR
SHNNKIQSTAS"
9683 /gene="BPTF"
2 2 GCCAGGCTGAGTGGCGCCCAAGACGCGGCTGAGCTCGCCAGGCGGCGGACGAGTACGC 61
|||||
215 GCCAGGCTGAGTGGCGCCCAAGACGCGGCTGAGCTCGCCAGGCGGCGGACGAGTACGC 274
|||||
62 GGAGGAAGCGGGG 121
|||||
275 GGAGGAAGCGGGG 334
|||||
122 GCGAGGAGCGGA 181
|||||
335 GCGAGGAGCGGA 394
|||||
182 CCACCG 241
|||||
395 CCACCG 454
|||||
242 TGGAGGAAGAGGAGGACATGCTGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301
|||||
455 TGGAGGAAGAGGAGGACATGCTGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 514
|||||
302 AGACCCAGGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 361
|||||
515 AGACCCAGGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 574
|||||
362 ATTATCCGAGGAGATGGAAGACGACGACGACGACGACGACGACGACGACGACGACGACG 421
|||||
575 ATTATCCGAGGAGATGGAAGACGACGACGACGACGACGACGACGACGACGACGACGACG 634
|||||
422 TCAGGAGCCATAGTACCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 481
|||||
635 TCAGGAGCCATAGTACCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 694
|||||
482 CTGCTTCTCTATATGGAAGAAAACACATCCGCGCCCTTGAATTTCCCAAGTCTCTCTG 541
|||||
695 CTCGTTCTCTATATGGAAGAAAACACATCCGCGCCCTTGAATTTCCCAAGTCTCTCTG 754
|||||
542 AGGATTTAATGGTGCCTTAATGAGCATATAATGATGTCATTGCCATTTACGAGGTACTGC 601
|||||
755 AGGATTTAATGGTGCCTTAATGAGCATATAATGATGTCATTGCCATTTACGAGGTACTGC 814
|||||
602 GGAATTTGGCAGCTGTTTGGATTAATCTCTTTTCCTTTGAGGACTTTTGTGCAAGCTC 661
|||||
815 GGAATTTGGCAGCTGTTTGGATTAATCTCTTTTCCTTTGAGGACTTTTGTGCAAGCTC 874
|||||
662 TGGTGAGCAGACAGCAGTGCACACTCATGCGAGAGATGCATGTTGTGCTTTTGAAGACAG 721
|||||
875 TGGTGAGCAGACAGCAGTGCACACTCATGCGAGAGATGCATGTTGTGCTTTTGAAGACAG 934
|||||
722 TTCTGCGTGAAGAAAGACACTTCCCAATACCTACCTTTGGACCTGCTGATCTGAAGATAGCG 781
|||||
935 TTCTGCGTGAAGAAAGACACTTCCCAATACCTACCTTTGGACCTGCTGATCTGAAGATAGCG 994
|||||
782 TTAATTCACACTGTTTTCATAGATGGATGAGTGCAGGAGAGGAGTCTCGGGGTGACT 841
|||||
995 TTAATTCACACTGTTTTCATAGATGGATGAGTGCAGGAGAGGAGTCTCGGGGTGACT 1054
|||||
842 GTGAGAGTGATAAGAGGTACCATCACGTTCTTCTTACCAAGAGGAGAGGAGTACCAT 901
|||||
1055 GTGAGAGTGATAAGAGGTACCATCACGTTCTTCTTACCAAGAGGAGAGGAGTACCAT 1114
|||||
902 ATGGACACAGTAGAAGAACAGATCAAAAGTTCTACAGTTTCTAGTCATCAGTTCTTACAA 961
|||||

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
source

gene
CDS

Db	1115	ATGACCACTAGAGAACAGATCAAAAGTTCTACAGTTCTAGTCGATCAGTTCTTACAA	1174	1918	-----	1917
QY	962	CAAATATTGCTCGAGAGGAATTGATGCTCTGAAGGGGTGATACAGTATGATGACCAATTGTA	1021	2042	AAAGAAATCTCATACACCTGTCTCTATTTCAGGAAGAGATAGTAGTGTATTTCACATCGG	2101
Db	1175	CAAATATTGCTCGAGAGGAATTGATGCTCTGAAGGGGTGATACAGTATGATGACCAATTGTA	1234	1918	-----	1936
QY	1022	GGTTTGTACAAAACCTTGGGGATTGCTTTGCTGTGTGAGACATGTTTCAGCAGTATACCAATT	1081	2102	AGAAGTCCACCGGGGAGCTAAGTGAATCTCCTCGAGCTGGAAAGAGAGCATCTGGGTCAA	2161
Db	1235	GGTTTGTACAAAACCTTGGGGATTGCTTTGCTGTGTGAGACATGTTTCAGCAGTATACCAATT	1294	1937	AGAAGTCCACCGGGGAGCTAAGTGAATCTCCTCGAGCTGGAAAGAGAGCATCTGGGTCAA	1996
QY	1082	TGGAATGTGTGAAGCCACCTCTTGAAGAGGTGTCAGAGGAGTGCAGAGGAGTGGCAGTCT	1141	2162	CTCGAATCATCACACGATTCGGAATCCAGATAGCAAACTTACTAGCTGAAGAGCCAGC	2221
Db	1295	TGGAATGTGTGAAGCCACCTCTTGAAGAGGTGTCAGAGGAGTGGCAGTCTGGAAGTCT	1354	1997	CTCGAATCATCACACGATTCGGAATCCAGATAGCAAACTTACTAGCTGAAGAGCCAGC	2056
QY	1142	GTGTAGCACACAAGGTGCTGTGTGTGACTGACTGTGTGCTGAAATCCAAAAATAAAC	1201	2222	AGGTGCAGCCGCTGCACATGAAGCAATAAATATTTAAGGAGGCAAGAGTACTGG	2281
Db	1355	GTGTAGCACACAAGGTGCTGTGTGTGACTGACTGTGTGCTGAAATCCAAAAATAAAC	1414	2057	AGGTGCAGCCGCTGCACATGAAGCAATAAATATTTAAGGAGGCAAGAGTACTGG	2116
QY	1202	CATATATTGCACATGAACCTATTGGATATGATAGAAAGTCGGAGGAATATCTGGTTCTTGA	1261	2282	TAGTTAACTCTCAAGGAGAAATTTTCAAGGAGGAGTTCAGGTTGAGACCCAAAAAGGAGTATCATGAAAG	2341
Db	1415	CATATATTGCACATGAACCTATTGGATATGATAGAAAGTCGGAGGAATATCTGGTTCTTGA	1474	2117	TAGTTAACTCTCAAGGAGAAATTTTCAAGGAGGAGTTCAGGTTGAGACCCAAAAAGGAGTATCATGAAAG	2176
QY	1262	ACCGAAGACTCATAATAGAGAGATACAGAAAATGAAAATGAAAAGAAATTTGGTATT	1321	2342	GAATATCAACAATTTTAAATTTGGTCAAGAGGAGTATCCGCTTACCACAATC	2401
Db	1475	ACCGAAGACTCATAATAGAGAGATACAGAAAATGAAAATGAAAAGAAATTTGGTATT	1534	2177	GAATATCAACAATTTTAAATTTGGTCAAGAGGAGTATCCGCTTACCACAATC	2236
QY	1322	ACAGCACAAAGGTCCAACTTGCAGAAATTAATGACTGCTAGACAAAGATTTATGGGAAG	1381	2402	AACTACTCCCACTTCAATTTGCTTGAATAAGCACCAGACAGAGAGACCATGATAAGA	2461
Db	1535	ACAGCACAAAGGTCCAACTTGCAGAAATTAATGACTGCTAGACAAAGATTTATGGGAAG	1594	2237	AACTACTCCCACTTCAATTTGCTTGAATAAGCACCAGACAGAGAGACCATGATAAGA	2296
QY	1382	CAGAACTCTGCAAAATTCAGAAAGAAATCGGTGAAGAAATCCACCACACATGGACATAA	1441	2462	GAAGGCATCTTGACACATAAGTTCTGCTGACTCCAGCAGGAGAGTTCAAAATGGAACGGTT	2521
Db	1595	CAGAACTCTGCAAAATTCAGAAAGAAATCGGTGAAGAAATCCACCACACATGGACATAA	1654	2297	GAAGGCATCTTGACACATAAGTTCTGCTGACTCCAGCAGGAGAGTTCAAAATGGAACGGTT	2356
QY	1442	CTGAAGACCTGACCAATAGGCTCGGGCAGTAACAATCCTTCTCGGGCAGCTAATG	1501	2522	CTGTCCATGGTCCAAAGTCTTACCATATCTACTCTGAGACTGACTATCACCACCAATTAG	2581
Db	1655	CTGAAGACCTGACCAATAGGCTCGGGCAGTAACAATCCTTCTCGGGCAGCTAATG	1714	2357	CTGTCCATGGTCCAAAGTCTTACCATATCTACTCTGAGACTGACTATCACCACCAATTAG	2416
QY	1502	AAGAAATTTTGAATCCATAAGAGCCAAAGGAGAGACATTTGATAATGTTAAAGCCACAG	1561	2582	AAACAACTATCCCTTCTCATCTTTCATCCCACTGGGCATCACATAGGCAAAATGGA	2641
Db	1715	AAGAAATTTTGAATCCATAAGAGCCAAAGGAGAGACATTTGATAATGTTAAAGCCACAG	1774	2417	AAACAACTATCCCTTCTCATCTTTCATCCCACTGGGCATCACATAGGCAAAATGGA	2476
QY	1562	AAGAAACAGAAAAGACAAGATGAGCTGAGAAATGCTCTAAAGATGCTGAGAAAAACA	1621	2642	TCAGGCAGTTCAGATGTGTAGCAAAACCAGAGAAATTTGCATTTGGCTTTAGCCATTTTGG	2701
Db	1775	AAGAAACAGAAAAGACAAGATGAGCTGAGAAATGCTCTAAAGATGCTGAGAAAAACA	1834	2477	TCAGGCAGTTCAGATGTGTAGCAAAACCAGAGAAATTTGCATTTGGCTTTAGCCATTTTGG	2536
QY	1622	GAGAAGATTTGAAGACCAGTCCCTTGAAGAGACAGTCCAGCAAAACACACAGATGATG	1681	2702	AGTGTCCAGTTAAACCAAGTTGTGATGCTACCAATATGGCGAGAAATTTTGGAGACATACCA	2761
Db	1835	GAGAAGATTTGAAGACCAGTCCCTTGAAGAGACAGTCCAGCAAAACACACAGATGATG	1894	2537	AGTGTCCAGTTAAACCAAGTTGTGATGCTACCAATATGGCGAGAAATTTTGGAGACATACCA	2596
QY	1682	ACCCTGAGCAAGGAAATCTGAGGAGCCCAACAGAAAGTTGGGGATAAAGTAACTCTGTGT	1741	2762	GGTTACACCGGATGACATCAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2821
Db	1895	ACCCTGAGCAAGGAAATCTGAG	1917	2597	GGTTACACCGGATGACATCAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2656
QY	1742	CAGCAAACTTGGCGAACAACAACAATAATGCAACTTCAGAGAGACTAGTCCCTCTGAGAG	1801	2822	AGAAACAG	2881
Db	1918	-----	1917	2657	AGAAACAG	2716
QY	1802	GGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCCCGATAGCAGCAACATGGCAGAGAGCA	1861	2882	TTAAGCATCAGGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTACAGAGATATGGTGGTT	2941
Db	1918	-----	1917	2717	TTAAGCATCAGGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTACAGAGATATGGTGGTT	2776
QY	1862	AGGTGSCATCTGAGCTCCCGGAGTGTGCCAGAGAAACCTAACAGACATGTGAGAGCA	1921	2942	GGAGCTGGATTAGTAAACTCATGTTTATAGGTTTGTCTCCTAAATTTGCCAGGCAATACTA	3001
Db	1918	-----	1917	2777	GGAGCTGGATTAGTAAACTCATGTTTATAGGTTTGTCTCCTAAATTTGCCAGGCAATACTA	2836
QY	1922	GTAACACTAGTGTACCACTACCTCCATCCAGCCTAATCTGAAAAACAGTAAACAGAGCA	1981	3002	ATGTGAATTACAGAAAAGTCTGTAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3061
Db	1918	-----	1917	2837	ATGTGAATTACAGAAAAGTCTGTAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2896
QY	1982	GTGAACATAATTTCTCCAGAGTGAATCTGCTAAGGCAGCTGATGATCTCTGAAATGGAG	2041	3062	AGTCACATAAAAAGAAATGTTACAGAGTCCAAAAAATAAATAATAGAGCCTGATCTCG	3121
				2897	AGTCAGATAAAAAGAAATGTTACAGAGTCCAAAAAATAAATAATAGAGCCTGATCTCTG	2956

QY 3122 AAAAGATGAGGTAAAGGTTTCAGATGCTGCCAAAAGAGCAGACCAAAAATGAAATGGATA 3181
 DB 2957 AAAAGATGAGGTAAAGGTTTCAGATGCTGCCAAAAGAGCAGACCAAAAATGAAATGGATA 3016
 QY 3182 TCTCAAGATTTACTGAGAAGAGCACCAGATCTGAAGGAGCTTTAGATTTCTGACAGTG 3241
 DB 3017 TCTCAAGATTTACTGAGAAGAGCACCAGATCTGAAGGAGCTTTAGATTTCTGACAGTG 3076
 QY 3242 ATAAACCTGCAAGGAAGAACCAATGGAAGTAGACGATGACATGAAACAGAGTCACATG 3301
 DB 3077 ATAAACCTGCAAGGAAGAACCAATGGAAGTAGACGATGACATGAAACAGAGTCACATG 3136
 QY 3302 TAAATTTGTCAGGAGAGTTCTCAAGTAGATGTTGGTCAATGTTAGTGGGTTTTTCATCTAA 3361
 DB 3137 TAAATTTGTCAGGAGAGTTCTCAAGTAGATGTTGGTCAATGTTAGTGGGTTTTTCATCTAA 3196
 QY 3362 GGACTAGTTACAAAAGAAAACAAAATCATCCAACTAGATGGACTTTCTTGAAGAGGAA 3421
 DB 3197 GGACTAGTTACAAAAGAAAACAAAATCATCCAACTAGATGGACTTTCTTGAAGAGGAA 3256
 QY 3422 TTAACAGTTTACACTGGAAGAAAACAGCGACTCGAAAANAATCAAGTTGGAGGGTGGAA 3481
 DB 3257 TTAACAGTTTACACTGGAAGAAAACAGCGACTCGAAAANAATCAAGTTGGAGGGTGGAA 3316
 QY 3482 TTAAGGGTATAGGAAGAGCTTCTACAAATCTTCAAAAATCTCTCTGAATCACAGTAA 3541
 DB 3317 TTAAGGGTATAGGAAGAGCTTCTACAAATCTTCAAAAATCTCTCTGAATCACAGTAA 3376
 QY 3542 TAACGAAAAGAAAAGAGGTGTCCAGAGTGAATCGATGAGACAAGACAGAGCCCAATG 3601
 DB 3377 TAACGAAAAGAAAAGAGGTGTCCAGAGTGAATCGATGAGACAAGACAGAGCCCAATG 3436
 QY 3602 CAATATGATCAACCTGAGACATGTTGAGGATGTTCAAGATGTTCAAGTTCCTCAGTTC 3661
 DB 3437 CAATATGATCAACCTGAGACATGTTGAGGATGTTCAAGATGTTCAAGTTCCTCAGTTC 3496
 QY 3662 TTAGAATGAGTGAATCTAGTGCATACCAACAAACAACTTTATCCAAAAGATCGAGTGTAG 3721
 DB 3497 TTAGAATGAGTGAATCTAGTGCATACCAACAAACAACTTTATCCAAAAGATCGAGTGTAG 3556
 QY 3722 ATGATGTCTCCATTCGGAGCCAGAACAAAATGTCGAAAACAAAATTCATTCAAAATG 3781
 DB 3557 ATGATGTCTCCATTCGGAGCCAGAACAAAATGTCGAAAACAAAATTCATTCAAAATG 3616
 QY 3782 ACATAGAGAAAAGTCTGACCTTGCCAGTGTAGAGGCCAGAACCCCACTAAGAGTAAAA 3841
 DB 3617 ACATAGAGAAAAGTCTGACCTTGCCAGTGTAGAGGCCAGAACCCCACTAAGAGTAAAA 3676
 QY 3842 CCAAGGAAATGATTTTTTCATTCGATGACTCTAAACTAGCCAGTGCAGATGATATGGTA 3901
 DB 3677 CCAAGGAAATGATTTTTTCATTCGATGACTCTAAACTAGCCAGTGCAGATGATATGGTA 3736
 QY 3902 CTTTGATCTGTGAAGACAAAACCGCTCATACAGGAGGAAGTGACACCATGTTTCTT 3961
 DB 3737 CTTTGATCTGTGAAGACAAAACCGCTCATACAGGAGGAAGTGACACCATGTTTCTT 3796
 QY 3962 CTTTCAAGAGTCTTTTACATTCATCAGTCCCTTAAAGTACCATGACAGAGATGCCAC 4021
 DB 3797 CTTTCAAGAGTCTTTTACATTCATCAGTCCCTTAAAGTACCATGACAGAGATGCCAC 3856
 QY 4022 CTCCTGTCAAGAGCAATGGACTTTGAAGGAAAACCTGGGATGTCGACTCTGAATCTAATAGA 4081
 DB 3857 CTCCTGTCAAGAGCAATGGACTTTGAAGGAAAACCTGGGATGTCGACTCTGAATCTAATAGA 3916
 QY 4082 CTTTGAANAATAGTTCTGTATCCGCTCTATTTCAGATAGCAGTGAAGAAGATATGATTG 4141
 DB 3917 CTTTGAANAATAGTTCTGTATCCGCTCTATTTCAGATAGCAGTGAAGAAGATATGATTG 3976
 QY 4142 TTCAGAAATAGCAATGAAGCAATTTCTGAACAGTTCTGAACTCCAGAACCAAGATGTTGAG 4201
 DB 3977 TTCAGAAATAGCAATGAAGCAATTTCTGAACAGTTCTGAACTCCAGAACCAAGATGTTGAG 4036

QY 4202 TCTTGAGCCGTTAAAGTGTGAGTTGGTTCTGTGAGTCCACTGGAAACTGTGAGGACA 4261
 DB 4037 TCTTGAGCCGTTAAAGTGTGAGTTGGTTCTGTGAGTCCACTGGAAACTGTGAGGACA 4096
 QY 4262 GCTGCCGCTCAAGGGGACTGAAGCAAAATGGTAAAAAACCAAGTCAGCAGAGAAGAAATAG 4321
 DB 4097 GCTGCCGCTCAAGGGGACTGAAGCAAAATGGTAAAAAACCAAGTCAGCAGAGAAGAAATAG 4156
 QY 4322 AGGAGAGACCACTTAATAATCTAGTGCATCAATAAAGCTAAAAATACCACTGCACAAA 4381
 DB 4157 AGGAGAGACCACTTAATAATCTAGTGCATCAATAAAGCTAAAAATACCACTGCACAAA 4216
 QY 4382 AGAATAATGAAATCGAGAGTCTGAAAAGAGAGCAGAGAACAAGTACATTTCAAAATA 4441
 DB 4217 AGAATAATGAAATCGAGAGTCTGAAAAGAGAGCAGAGAACAAGTACATTTCAAAATA 4276
 QY 4442 ATGGAAGAATATAAACCACCAAAAATATTTGAAAGGTGAATGCTTGAAGAAATTTCTG 4501
 DB 4277 ATGGAAGAATATAAACCACCAAAAATATATTTGAAAGGTGAATGCTTGAAGAAATTTCTG 4336
 QY 4502 AGAGTAGAGTGTAGTAACTGTTTGAACCAAGGTTAATATAATAAATAAATAATCC 4561
 DB 4337 AGAGTAGAGTGTAGTAACTGTTTGAACCAAGGTTAATATAATAAATAAATAATCC 4396
 QY 4562 CTGAGAAATGATATTAATCATTTGACTGTTAAAGAAATCTGCTATAAGGCCATTCATTAATG 4621
 DB 4397 CTGAGAAATGATATTAATCATTTGACTGTTAAAGAAATCTGCTATAAGGCCATTCATTAATG 4456
 QY 4622 GTGATGTCATCATGGAAGATTTTAAATGAAAGAACAGCTCCGAAACAAAATGCAATTCG 4681
 DB 4457 GTGATGTCATCATGGAAGATTTTAAATGAAAGAACAGCTCCGAAACAAAATGCAATTCG 4516
 QY 4682 TGAGTTCTTCAGATGCTGAAGGTAACCTACCGAGATAGCTTTGAGACCCCTGCCATCAACA 4741
 DB 4517 TGAGTTCTTCAGATGCTGAAGGTAACCTACCGAGATAGCTTTGAGACCCCTGCCATCAACA 4576
 QY 4742 AAGAGTCTGACGTACACAGACAGCACCACACCTCAGCATCTTGTCCAGAAAGCAATTCAG 4801
 DB 4577 AAGAGTCTGACGTACACAGACAGCACCACACCTCAGCATCTTGTCCAGAAAGCAATTCAG 4636
 QY 4802 TTAATCAGGTAGAGATATGGAATAGAAACCTCAGAACTTAAGAAAGTAAAGTACTTCATC 4861
 DB 4637 TTAATCAGGTAGAGATATGGAATAGAAACCTCAGAACTTAAGAAAGTAAAGTACTTCATC 4596
 QY 4862 CTATTACTTCTGAAGAGGAATCTAACTCAGTAATGACTTTTATGATGAAAATGGTCTGC 4921
 DB 4697 CTATTACTTCTGAAGAGGAATCTAACTCAGTAATGACTTTTATGATGAAAATGGTCTGC 4756
 QY 4922 CCATCAACAAAATGAAAATGTCATGGAAGAAATCTTAAAGAAAACCCGTCATCACAGAAG 4981
 DB 4757 CCATCAACAAAATGAAAATGTCATGGAAGAAATCTTAAAGAAAACCCGTCATCACAGAAG 4816
 QY 4982 TCACACGATGACCTCCACAGTGGCCACAGAAATCAAAAACCTGATCAAGGTAGAAAAG 5041
 DB 4817 TCACACGATGACCTCCACAGTGGCCACAGAAATCAAAAACCTGATCAAGGTAGAAAAG 4876
 QY 5042 GCGATTAAGCAAACTGTGGTTTCTCCACAGAAAATTTGCAAAAATCCACTGTGCACACCA 5101
 DB 4877 GCGATTAAGCAAACTGTGGTTTCTCCACAGAAAATTTGCAAAAATCCACTGTGCACACCA 4936
 QY 5102 CCAGTACAAACAGTACCAGGCTTTCCACACCCCTCCACAGGCGGAGTGTGGACATCATCT 5161
 DB 4937 CCAGTACAAACAGTACCAGGCTTTCCACACCCCTCCACAGGCGGAGTGTGGACATCATCT 4996
 QY 5162 CTGTAAGGAGCAGAGCAAAACCGTGTCCACAGCAGAGTACAGAGCTCCCTGACACCA 5221
 DB 4997 CTGTAAGGAGCAGAGCAAAACCGTGTCCACAGCAGAGTACAGAGCTCCCTGACACCA 5056
 QY 5222 CGGAGGACACACTGGTTACATCTATGACTGTGAGCAAGAGTATTTCCACAGCAGACAAAG 5281
 DB 5057 CGGAGGACACACTGGTTACATCTATGACTGTGAGCAAGAGTATTTCCACAGCAGACAAAG 5116
 QY 5282 TGAACACTGATGAAATTTTCAAGACCAAAAGAGACTGCTTTCAGGTACAGCTTCGCATCCT 5341

Db 5117 TGAACAGTGAATAATTTCAAGACCAAGAAAGACTCGTTTCAGGTACAGCTCTGCCATCCT 5176
QY 5342 ATAGAAAAATTTGTACCAAGAGACCAAGAAAGAGCAATTTTGTGTTTGCCTTAATGATGACT 5401
Db 5177 ATAGAAAAATTTGTACCAAGAGACCAAGAAAGAGCAATTTTGTGTTTGCCTTAATGATGACT 5236
QY 5402 TAAAAAGTTGGCCCGAAAGGAGGAATCCGAGAGGTCCCTTATTTTAATTAACAATGCAA 5461
Db 5237 TAAAAAGTTGGCCCGAAAGGAGGAATCCGAGAGGTCCCTTATTTTAATTAACAATGCAA 5296
QY 5462 AACCTGCTTTGGATATATGGCCATATCCCTTCCTTAGACCGACCTTTGGCATCAGTTGGA 5521
Db 5297 AACCTGCTTTGGATATATGGCCATATCCCTTCCTTAGACCGACCTTTGGCATCAGTTGGA 5356
QY 5522 GGTATAGACTTCAGACAGTAAAGTCTTAGCTGGAGTGAGCCTGATGTTACGGTTACTGT 5581
Db 5357 GGTATAGACTTCAGACAGTAAAGTCTTAGCTGGAGTGAGCCTGATGTTACGGTTACTGT 5416
QY 5582 GGGCAAGTTTGAGATGGGATGATATGGCGGCGCAAGGTTCCCTCAGGAGGAGGAGTACAC 5641
Db 5417 GGGCAAGTTTGAGATGGGATGATATGGCGGCGCAAGGTTCCCTCAGGAGGAGGAGTACAC 5476
QY 5642 GGCAGAAACATCCGAAACTGAAATCACAACACAGAAATATTAAGAGGAGAGATGTTG 5701
Db 5477 GGCAGAAACATCCGAAACTGAAATCACAACACAGAAATATTAAGAGGAGAGATGTTG 5536
QY 5702 GTCCCTTATGGCATTCGATTTGAATATTGTATCAGGAAATCATTTCTCCATTGGAGTTC 5761
Db 5537 GTCCCTTATGGCATTCGATTTGAATATTGTATCAGGAAATCATTTCTCCATTGGAGTTC 5596
QY 5762 CAGAAACACCAAAAGAAACGCTACACCTCAGAGGAAAGCCCTTCGATCAAGTGCACTGC 5821
Db 5597 CAGAAACACCAAAAGAAACGCTACACCTCAGAGGAAAGCCCTTCGATCAAGTGCACTGC 5656
QY 5822 GGCRAAGAGACGAGAAAGCCCAAGCAACTGGCCCTGTTATTTAAAGAACTGGGTAG 5881
Db 5657 GGCRAAGAGACGAGAAAGCCCAAGCAACTGGCCCTGTTATTTAAAGAACTGGGTAG 5716
QY 5882 CAGAGAAAGAACTGGAATTTGTGGAGATCAGGGCATTTGCTGAGAGAGTGGAGAAAGAA 5941
Db 5717 CAGAGAAAGAACTGGAATTTGTGGAGATCAGGGCATTTGCTGAGAGAGTGGAGAAAGAA 5776
QY 5942 AGGCAACAGCTGTGAGCAACAGGCTTAAGAAAGCACTGGAGCAGCAGAGAGCGACAGTGA 6001
Db 5777 AGGCAACAGCTGTGAGCAACAGGCTTAAGAAAGCACTGGAGCAGCAGAGAGCGACAGTGA 5836
QY 6002 TTGCAACTTCCACTACTTCCCAACAGCAGTACACAGCAGCAGCAGCAGCAGCAGCAGCAG 6061
Db 5837 TTGCAACTTCCACTACTTCCCAACAGCAGTACACAGCAGCAGCAGCAGCAGCAGCAGCAG 5896
QY 6062 AAGTTATGTTGGCCCGCCCATTAAGTGGCTCAGTTTACAACTGGAACCAAAATGGTACTA 6121
Db 5897 AAGTTATGTTGGCCCGCCCATTAAGTGGCTCAGTTTACAACTGGAACCAAAATGGTACTA 5956
QY 6122 CTAAGTTGATCTCCAGCTACAGTAACTATCCCAACAAAGAACTTTCATCAAACT 6181
Db 5957 CTAAGTTGATCTCCAGCTACAGTAACTATCCCAACAAAGAACTTTCATCAAACT 6016
QY 6182 TTGCTACATGGGTAAAGCAAGCCAGTCAAAATTCAGGCGTTGTTCAAGTACAGCAGAAAG 6241
Db 6017 TTGCTACATGGGTAAAGCAAGCCAGTCAAAATTCAGGCGTTGTTCAAGTACAGCAGAAAG 6076
QY 6242 TCCTGGGTATCATTCATCAAGTACAGGTACCAGTCAGCAAACTTTACTTTCATTCAGC 6301
Db 6077 TCCTGGGTATCATTCATCAAGTACAGGTACCAGTCAGCAAACTTTACTTTCATTCAGC 6136
QY 6302 CCAGGACAGCAAGTACAAATTAGGCCCAATACCTCAGGCTCTGGAGAAACCAAGCA 6361
Db 6137 CCAGGACAGCAAGTACAAATTAGGCCCAATACCTCAGGCTCTGGAGAAACCAAGCA 6196
QY 6362 ATTCAGAAATATCAGAGGCGCTCAGATTGCGCCCTGGTATGACCGGTGATTAGAACACCA 6421
Db

Db 6197 ATTCAACAAGTAATACAGAGGCGCTCAGATTGCGCCCTGGTATGACCGGTGATTAGAACACCA 6356
QY 6422 TCCAAACAGTCAACACTAGGAAAGGCAATTAATTCGAACACACTGTGATGGTACAGCCAGGTG 6481
Db 6257 TCCAAACAGTCAACACTAGGAAAGGCAATTAATTCGAACACACTGTGATGGTACAGCCAGGTG 6316
QY 6482 CTCCTCAGCAAGTGATGACTCAAAATCATCAGGGGGGAGCCTGTCTCCACTGCAAGTCTCCG 6541
Db 6317 CTCCTCAGCAAGTGATGACTCAAAATCATCAGGGGGGAGCCTGTCTCCACTGCAAGTCTCCG 6376
QY 6542 CCCTTAACACGGTTTCTCTAACACCTTGGCGAGAAAAGCTTAACCTTCAGCAACCTCCACTT 6601
Db 6377 CCCTTAACACGGTTTCTCTAACACCTTGGCGAGAAAAGCTTAACCTTCAGCAACCTCCACTT 6436
QY 6602 CAAATATACAGTCTTCAGGCTCACAACCCCTCGCCCAACAAAGGACAAAGTGAAGCTCA 6661
Db 6437 CAAATATACAGTCTTCAGGCTCACAACCCCTCGCCCAACAAAGGACAAAGTGAAGCTCA 6496
QY 6662 CCATGGCTCAACTTACTCAGTTAAACACAGGGCCACGGTGGCAATCAAGGTTTGACAGTAG 6721
Db 6497 CCATGGCTCAACTTACTCAGTTAAACACAGGGCCACGGTGGCAATCAAGGTTTGACAGTAG 6556
QY 6722 TAATTCAAGSACAAGCTCAAACTACTGGACAGTTGCAGTTGATACCTCAAGGGGTGACTG 6781
Db 6557 TAATTCAAGSACAAGCTCAAACTACTGGACAGTTGCAGTTGATACCTCAAGGGGTGACTG 6616
QY 6782 TACTCCAGGCGCCAGGCGCAGCAGCTAATGCAAGCTGCAATGCCAAATGGTACTGTTTCAGC 6841
Db 6617 TACTCCAGGCGCCAGGCGCAGCAGCTAATGCAAGCTGCAATGCCAAATGGTACTGTTTCAGC 6676
QY 6842 GATTCCTCTTTTACCCCATTTGGCAACAAACAGCCACCAGCAGCAGCAGCAGCAGCAGCAGT 6901
Db 6677 GATTCCTCTTTTACCCCATTTGGCAACAAACAGCCACCAGCAGCAGCAGCAGCAGCAGCAGT 6736
QY 6902 TTTCCACGACAGCAGCAGGTACAGGTGAACAAGGACAGAGTAACCTGTCACCCACAGATGC 6961
Db 6737 TTTCCACGACAGCAGCAGGTACAGGTGAACAAGGACAGAGTAACCTGTCACCCACAGATGC 6796
QY 6962 AGGTACATCAAGACAAACCCCTGCCACAGCTCAGTCATCAAGTGGGGTCCAGCAAAAG 7021
Db 6797 AGGTACATCAAGACAAACCCCTGCCACAGCTCAGTCATCAAGTGGGGTCCAGCAAAAG 6856
QY 7022 CCCAGCCACAGCTGCTCAGCCTTACGCTGGCCCGCAGCCCAACCCAGCCCGCCAGTCCC 7081
Db 6857 CCCAGCCACAGCTGCTCAGCCTTACGCTGGCCCGCAGCCCAACCCAGCCCGCCAGTCCC 6916
QY 7082 CAGCTCAGCTGAAAGTTTCAGACTCAGCTGAAAGTTTCAGACCCCAACCACTGTTTCATCCC 7141
Db 6917 CAGCTCAGCTGAAAGTTTCAGACTCAGCTGAAAGTTTCAGACCCCAACCACTGTTTCATCCC 6976
QY 7142 ATGTCCCTTCTGAAGCACAACCCAGCAGCAGTCATCCAAAGCCCAAGTTGACAGCAG 7201
Db 6977 ATGTCCCTTCTGAAGCACAACCCAGCAGCAGTCATCCAAAGCCCAAGTTGACAGCAG 7036
QY 7202 AGTCTCAGCCTCAAGTAAATGTCACAGGACAGTCTCCTGTTGTTGTCGTCGTCGTCGTCGTC 7261
Db 7037 AGTCTCAGCCTCAAGTAAATGTCACAGGACAGTCTCCTGTTGTTGTCGTCGTCGTCGTCGTC 7096
QY 7262 AGACTCGAATACGTCATCAACTCCATCCCAACTGTCCTCTGAGCAACCAATCCAGGTTTC 7321
Db 7097 AGACTCGAATACGTCATCAACTCCATCCCAACTGTCCTCTGAGCAACCAATCCAGGTTTC 7156
QY 7322 AGACTCAACCTCACAACCGATTTCCAATTCACACCATATCATCTCTTCAGATACCTTCCC 7381
Db 7157 AGACTCAACCTCACAACCGATTTCCAATTCACACCATATCATCTCTTCAGATACCTTCCC 7216
QY 7382 AAGGCAGCCACAGTCACAAACCCAGGTACAGTCTTCAACTCAAACTCTTTTCATTCAGGAC 7441
Db 7217 AAGGCAGCCACAGTCACAAACCCAGGTACAGTCTTCAACTCAAACTCTTTTCATTCAGGAC 7276
QY 7442 AACTTTTAATCAAGTTAGTGTTCATCCCATCCCGTCTCCTCAGCTACAAATACAGCAGC 7501
Db 7277 AACTTTTAATCAAGTTAGTGTTCATCCCATCCCGTCTCCTCAGCTACAAATACAGCAGC 7336

QY	7502	CACAGCCCCAAGTCATTGCTGCTGCCTCAGCTGCAACAAACAAGTCCAGGTTCTCTCTCAGA	7561	QY	8581	CAGTCCAGTCAACAGAGGATGCCATGACAGTGTCTCAGCCACTAACAGAGAAAGGATTAT	8640
DB	7337	CACAGCCCCAAGTCATTGCTGCTGCCTCAGCTGCAACAAACAAGTCCAGGTTCTCTCTCAGA	7396	DB	8416	CAGTCCAGTCAACAGAGGATGCCATGACAGTGTCTCAGCCACTAACAGAGAAAGGATTAT	8475
QY	7562	TCCAGTCACAGGTTGTGGCTCAGATACAGGCTCAGCAAAAGTGGTGTGCCCCAGCAAAATCA	7621	QY	8641	GAGGGTTGAAGAGGGTGTCTCGTTCCCTTACAGGCCCAAAAGATGAGTGGCTTCCCTT	8700
DB	7397	TCCAGTCACAGGTTGTGGCTCAGATACAGGCTCAGCAAAAGTGGTGTGCCCCAGCAAAATCA	7456	DB	8476	GAGGGTTGAAGAGGGTGTCTCGTTCCCTTACAGGCCCAAAAGATGAGTGGCTTCCCTT	8535
QY	7622	AACCTCCAGTTACCTATCCAAATTCAGAAAGCAGTGTGTGACAGCTCACCGATTCAGA	7681	QY	8701	GAACCATAGACCTAATGATGACACAGATTAATATGTTGTTTAAAGAACCTATGAC	8760
DB	7457	AACCTCCAGTTACCTATCCAAATTCAGAAAGCAGTGTGTGACAGCTCACCGATTCAGA	7516	DB	8536	GAACCATAGACCTAATGATGACACAGATTAATATGTTGTTTAAAGAACCTATGAC	8595
QY	7692	ATGTGGTTACAGTCGAGCGCCAGTGTGCAAGAGCAGTTGCAAGGGTTTCACCAACTCA	7741	QY	8761	CTTGCCACCATGGAAGAAGAGTACAAAGACCATATATGAAAAGCTGACGGAATTTGTG	8820
DB	7517	ATGTGGTTACAGTCGAGCGCCAGTGTGCAAGAGCAGTTGCAAGGGTTTCACCAACTCA	7576	DB	8596	CTTGCCACCATGGAAGAAGAGTACAAAGACCATATATGAAAAGCTGACGGAATTTGTG	8655
QY	7742	GGGATCAGCAGCAAAAGAAAGAACAGCAAGATAGAAATTAACGTTGAACACACACCCCTCC	7800	QY	8821	GCAGATATCACCAAAATTTTGTATTAACCTGCTTACTACAAATCCAAAGTGACTCCCAATTT	8880
DB	7577	GGGATCAGCAGCAAAAGAAAGAACAGCAAGATAGAAATTAACGTTGAACACACACCCCTCC	7636	DB	8656	GCAGATATCACCAAAATTTTGTATTAACCTGCTTACTACAAATCCAAAGTGACTCCCAATTT	8715
QY	7801	AAGCTTCTAATCAAAGTTGAATCATTCAGAAACAGTGTGTGATGAAGCATAATGCTGTA	7860	QY	8881	TACCACTGTGCAGAAAGTCTCGAATCATTTTGTACAGAAATTTGAAAGGCTTCAAGCT	8940
DB	7637	AAGCTTCTAATCAAAG-TGAATCATTCAGAAACAGTGTGTGATGAAGCATAATGCTGTA	7695	DB	8716	TACCACTGTGCAGAAAGTCTCGAATCATTTTGTACAGAAATTTGAAAGGCTTCAAGCT	8775
QY	7861	ATAGAACATTTAAACAGAAAGAGAGATGACTCCAGCTGAAAGAGAGAGAAATCAAAGA	7920	QY	8941	AGCAGTCTCATTAACAAACAACTGCAGTCTACAGTCTTCTTAAAGTTACAGCTGTTAACT	9000
DB	7696	ATAGAACATTTAAACAGAAAGAGAGATGACTCCAGCTGAAAGAGAGAGAAATCAAAGA	7755	DB	8776	AGCAGTCTCATTAACAAACAACTGCAGTCTACAGTCTTCTTAAAGTTACAGCTGTTAACT	8835
QY	7921	ATGATTTGCTGTAAACAGTGTGATGAAGTATATTTTGGATAGATAGATAGAAAGAAAA	7980	QY	9001	AACATAAAACACAGCAAGAACTCTGGTGTCTGCTGAACCTATTTTAAATTAAGAGCGAGATGT	9060
DB	7756	ATGATTTGCTGTAAACAGTGTGATGAAGTATATTTTGGATAGATAGATAGAAAGAAAA	7815	DB	8836	AACATAAAACACAGCAAGAACTCTGGTGTCTGCTGAACCTATTTTAAATTAAGAGCGAGATGT	8895
QY	7981	CAGGACGCAAAACGAAAGCGTGAAGAGAGTGTGAGAGAGAGAGTGTGAGAGAGAGAGT	8040	QY	9061	TTTTAGTCAGGCTATCCTGACAGACTGTGACCTTAAACTTCGTTTTTATTTGTTGTTAAACAG	9120
DB	7816	CAGGACGCAAAACGAAAGCGTGAAGAGAGTGTGAGAGAGAGAGTGTGAGAGAGAGAGT	7875	DB	8896	TTTTAGTCAGGCTATCCTGACAGACTGTGACCTTAAACTTCGTTTTTATTTGTTGTTAAACAG	8955
QY	8041	GCACCTTAAGCTGTACGCTCTGCTCTTCAAGCACAAGAGAGCTCAGAGCGGAGATCCTG	8100	QY	9121	TCCAAATATATCTTGGCCAAATTTTGTCCAAACGAGCAAGAAAAAGCAAAAGTCAACGACA	9180
DB	7876	GCACCTTAAGCTGTACGCTCTGCTCTTCAAGCACAAGAGAGCTCAGAGCGGAGATCCTG	7935	DB	8956	TCCAAATATATCTTGGCCAAATTTTGTCCAAACGAGCAAGAAAAAGCAAAAGTCAACGACA	9015
QY	8101	AGAAGAGAGACTCTCTGACAAAGATCTGCAAAATGAAGTGCAGAGAGAGCTGAAGAGA	8160	QY	9181	CCATTAATCTTCAAGATCAGATGGTGTCTTACTATTTGGCAGAGAGGAGAAAACTTTGTT	9240
DB	7936	AGAAGAGAGACTCTCTGACAAAGATCTGCAAAATGAAGTGCAGAGAGAGCTGAAGAGA	7995	DB	9016	CCATTAATCTTCAAGATCAGATGGTGTCTTACTATTTGGCAGAGAGGAGAAAACTTTGTT	9075
QY	8161	GACCTGAAATTAAGAAAGAAAAAGACCTGTAGTGTGGCTCAGGCCACAGCAGTACGT	8220	QY	9241	TATTTGAAAAAAGAAAAAGAAAGCAAGAAAAAGATATCTATGGGGTCAAGTGAACCT	9300
DB	7996	GACCTGAAATTAAGAAAGAAAAAGACCTGTAGTGTGGCTCAGGCCACAGCAGTACGT	8055	DB	9076	TATTTGAAAAAAGAAAAAGAAAGCAAGAAAAAGATATCTATGGGGTCAAGTGAACCT	9135
QY	8221	GCACCTGCCCCCAGTGACACAGTCTCTTCCAGCCCTCCAGCCCTCCAGCTTCACCT	8280	QY	9301	CCATGGAAATGCCACGCTCTGCTCTTCAAGTGAAGAGCTGGTTAGAGTCTCAGAGAAAC	9360
DB	8056	GCACCTGCCCCCAGTGACACAGTCTCTTCCAGCCCTCCAGCCCTCCAGCTTCACCT	8115	DB	9136	CCATGGAAATGCCACGCTCTGCTCTTCAAGTGAAGAGCTGGTTAGAGTCTCAGAGAAAC	9195
QY	8281	CCCCCTCCACCTGGTGTGCAACACAGCGCTCTCTCCACGCCACCTTACCTTGTGCT	8340	QY	9361	TTTTGACTGTATTTATTTATTTGTTGCAAAAAGACGCTTTTTTATTTGCTGCCCTCATTTG	9420
DB	8116	CCCCCTCCACCTGGTGTGCAACACAGCGCTCTCTCCACGCCACCTTACCTTGTGCT	8175	DB	9196	TTTTGACTGTATTTATTTATTTGTTGCAAAAAGACGCTTTTTTATTTGCTGCCCTCATTTG	9255
QY	8341	TCCAGAGAGAGCGGGAAGAGGAAAAAGACTCCAGCTCAAGTCCAAAGTCCAAAGAAAA	8400	QY	9421	TCAGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTACATATAATATATCTGTATCTT	9480
DB	8176	TCCAGAGAGAGCGGGAAGAGGAAAAAGACTCCAGCTCAAGTCCAAAGTCCAAAGAAAA	8235	DB	9256	TCAGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTACATATAATATATCTGTATCTT	9315
QY	8401	ATGATCTCTACTACCTCAAGGAACTAAGAGGACACAAAGCTTTTACTGTATCTGTAA	8460	QY	9481	ATCATGATTCCTGTAGGTAAAGTACAGAGCAGCTCTAGATGCTTTCTTTCTCTATGAA	9540
DB	8236	ATGATCTCTACTACCTCAAGGAACTAAGAGGACACAAAGCTTTTACTGTATCTGTAA	8295	DB	9316	ATCATGATTCCTGTAGGTAAAGTACAGAGCAGCTCTAGATGCTTTCTTTCTCTATGAA	9375
QY	8461	ACGCTTATCATGAATCTAAATTTTATTTGGCTGTGATCGGTTGATGAGTATGTTACCAT	8520	QY	9541	AGGAGCTCTGTATGATACATGTGCACACACACAACTGGGAATCAACAATGAGTTTAT	9600
DB	8296	ACGCTTATCATGAATCTAAATTTTATTTGGCTGTGATCGGTTGATGAGTATGTTACCAT	8355	DB	9376	AGGAGCTCTGTATGATACATGTGCACACACACAACTGGGAATCAACAATGAGTTTAT	9435
QY	8521	GGGCGCTGCGTTGGCATCTTTCGCAAGTGTAGGAGAGCTCATTTGATGAGTATGCTGTCA	8580	QY	9601	GTTTCATGTGTAGTATTAATTAAGCTTGCATAAAGGTTGGGCTTAAGTGTCTTGGGCTAC	9660
DB	8356	GGGCGCTGCGTTGGCATCTTTCGCAAGTGTAGGAGAGCTCATTTGATGAGTATGCTGTCA	8415	DB	9436	GTTTCATGTGTAGTATTAATTAAGCTTGCATAAAGGTTGGGCTTAAGTGTCTTGGGCTAC	9495
				QY	9661	AGACTCTGTTGCTTGAATATAACAGTACAAATTTGTTCAATTTACTCTGCACACAGGCTAAAG	9720

Db	9496	AGACTCTGGTTCCTTGAATATACACGATACAAATTTGTCATTAATCTCTGCACACGCTAAAG	9555	
Qy	9721	TGAGTAAATCTATTTCAGAGTATCTGTTGTGTAACATATTGTCAGATTTCTAAATTTTTT	9780	
Db	9556	TGAGTAAATCTATTTCAGAGTATCTGTTGTGTAACATATTGTCAGATTTCTAAATTTTTT	9615	
Qy	9781	CTTTTGTATTAAATTCACATCTGGATGTATATGAACACAAATAAATGGAGATAATTTTT	9840	
Db	9616	CTTTTGTATTAAATTCACATCTGGATGTATATGAACACAAATAAATGGAGATAATTTTT	9675	
Qy	9841	CTCCACAAAAA	9865	
Db	9676	CTCCACAAAAA	9700	
RESULT 2				
AY282495				
LOCUS	AY282495	8295 bp	mRNA	linear
DEFINITION	Homo sapiens bromodomain PHD finger transcription factor mRNA, complete cds.			
ACCESSION	AY282495			
VERSION	AY282495.1	GI:31322941		
KEYWORDS				
SOURCE				
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 8295) Barak,O., Lazzaro,M., Lane,W., Speicher,D., Picketts,D. and Shiekhhattar,R.			
TITLE	Isolation of human NURF			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 8295) Barak,O. and Shiekhhattar,R.			
AUTHORS	Direct Submission			
TITLE	Submitted (23-APR-2003) Gene Expression and Regulation, Wistar Institute, 3601 Spruce-St., Philadelphia, PA 19104, USA			
JOURNAL	Location/Qualifiers			
FEATURES	1. 8295 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /chromosome="17" /map="17q23" /cell_line="HEK293" 1. 8295 /note="BPTF; large subunit of the Homo sapiens nucleosome remodeling factor (NURF) complex containing BPTF, SNF2L, and RBAP46/48" /codon_start=1 /product="bromodomain PHD finger transcription factor" /translation="MYSEEEEDGDAAEETQDSEDEDEMEEDDDSDYPEEMEDDD DASYCTESFRSHYSTPTGPRKRPVHRPSPILEEKDIPLPEPKSEEDLMPNE HMINVTIYELVNFNGTGLRLSPERFEDCAALVSOEOKTFLMAEHVILKAVLRRED TSNITGPDADLKDVNSTLYFLIDGWTPEVLRVYCESDKYHVLPYQEAEDPYGPG ENKIKVLFYLDVDFLTNTAREELMSEGVIQYDDHCRVCHKGLDLCCECTSCVSHLE CVAPPLEEVEDNEWQCEVCAHVPGVTDCAEIQKNKPYRIHEPIGYDKRSRKYWFL NRRLLITTEEDTENENKKIWIYSTKVQLAELIDCLDQIWEAECLKILEEMREEIHRHM DITPDIETNARGSKNGSLAANEEILIESIRAKKGDIDNVKSPETEKNDETENDSKS AKNRREFEOLSKSDDKTDPDDPEQGSREPTVEGDKNGSVSANLGDNTNATNSE NLENSRSELNSQSSEKAAADDPENGERSHTPVIOEIVGDFXSEKSGELSES PGAGKASGASTRIITRLRPNDKSLQSKQAAAAAHEANKLFKEGKEVGLVNSQGEI SRLSTKEVIMKGINNYFKLQEGKRYVYHNQYTSNFSALNKHQHRDHDKRRHLAH KFCFLTAPGKFWKNSVHGSKVLITISLRLITOLENNIPSSFLHPNWSHRANRWIAK OMCSPREFALALAILECAVKPVMLPIWRESIGHTRLHRMTSIEREEKVKYKKEK QEEERTMQATWVYTFPVFKVQWVKQGEYRVTGIGGWSWISKTHVRFVPLPGNT NYNRYKSLGTNNMDENSDKRSKPPKIKIEPDSKEDVKSGDAAGADQNE MLDSIKTEKKDQVKKLLSDSKPKEPMEYDDMDKTESHVNCQESQVDVYVNSE GFHLRTYKSKTKSSKLDGLLERRIKOFTLEEKORLEKIKLEGGIKIGKTSNTNSK			
CDS				


```
QY 7399 CAACCCAGGTACAGTCTTCAACTCAAACTCTTTCATCAGGACAAACTTTAAATCAAGTT 7458
Db 7141 CAACCC----- 7146
QY 7459 AGTGTTCATCCCATCCGTCCTCAGCTACAAATACAGCAGCCACAGCCCAAGTCATT 7518
Db 7147 ----- 7146
QY 7519 GCTGTGCCTCAGCTGCAACAACAAGTCCAGGTTCTCTCTCAGATCCAGTCAAGGTTGTG 7578
Db 7147 ----- 7146
QY 7579 GCTCAGATACAGGCTCAGCAAAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATC 7638
Db 7147 ----- 7146
QY 7639 CAAATTCAGCAAGCAGTCTGTGCAGACTCACCAGATTCAGAAATGTGTTACAGTCGAG 7698
Db 7147 ----- 7146
QY 7699 GCAGCCAGTGTGCAAGAGCAGTTGCAAAAGGTTTCAGCAACTCAGGGATCAGCAGCAAAAG 7758
Db 7147 ----- 7146
QY 7759 AAGAAACAGCAACAGATAGAAATTAACGTGAACACACCCCTCCAAAGCTTCTAATCAAAAGTT 7818
Db 7147 ----- 7146
QY 7819 GAAATCATTCAGAAACAGTGGTATGAGCATATGCTGTAATAGAACATTTAAACAG 7878
Db 7147 -----CAGTGGTATGAGCATAATGCTGTAATAGAACATTTAAACAG 7191
QY 7879 AAAAAAGCATGACTCCAGCTGAAAGAGAGAGAAATCAAGAAATGATTGCTGTAAACAG 7938
Db 7192 AAAAGAGCATGACTCCAGCTGAAGAGAGAGAGATCAAGAAATGATTGCTGTAAACAG 7251
QY 7939 GTGATGAAGTATATTTTGGATAAGATAGATAAAGAGAAACACAGCGACGCAAAAAACGG 7998
Db 7252 GTGATGAAGTATATTTTGGATAAGATMGATAAAGAGAAACACAGCGACGCAAAAAACGG 7311
QY 7999 AAGCGTGAAGAGTGTGGAGCAGAAACGTACGACGAGATGCTGCACTAAAGCTGTCAGCT 8058
Db 7312 AAGCGTGAAGAGTGTGGAGCAGAAACGTACGACGAGATGCTGCACTAAAGCTGTCAGCT 7371
QY 8059 CTGCTCTTCAAGCAAAAGAGCAGCTCAGAGCCGAGATCCTGAAGAGAGAGACACTCCTG 8118
Db 7372 CTGCTCTTCAAGCAAAAGAGCAGCTCAGAGCCGAGATCCTGAAGAGAGAGACACTCCTG 7431
QY 8119 GACAAGGATCTGCAAAATTTGAAGTGCAGGAAGAGCTGAAGAGAGACCTGAAAATTAAGAAA 8178
Db 7432 GACAAGGATCTGCAAAATTTGAAGTGCAGGAAGAGCTGAAGAGAGACCTGAAAATTAAGAAA 7491
QY 8179 GAAAAGACCTGATGCACTGGCTCAGGCCACAGCAGTAGCTGCACCTGCCCTCCACCTGGTGTG 8238
Db 7492 GAAAAGACCTGATGCACTGGCTCAGGCCACAGCAGTAGCTGCACCTGCCCTCCACCTGGTGTG 7551
QY 8239 ACACGAGTCTTCCAGCCCTCCAGCCCTCCACCTTCCACCTCCACCTCCACCTGGTGTG 8298
Db 7552 ACACGAGTCTTCCAGCCCTCCAGCCCTCCACCTTCCACCTCCACCTCCACCTGGTGTG 7611
QY 8299 CAACACAGAGGCTTCTGTCCAGCCCTCCACCTTACCTGTGCTTCCAGAGAGGAGCGG 8358
Db 7612 CAACACAGAGGCTTCTGTCCAGCCCTCCACCTTACCTGTGCTTCCAGAGAGGAGCGG 7671
QY 8359 GAAGAGGAAAGACTCCAGCTCAAGTCCAGAGAAAAAGAAATGATCTCTACTACCTCA 8418
Db 7672 GAAGAGGAAAGACTCCAGCTCAAGTCCAGAGAAAAAGAAATGATCTCTACTACCTCA 7731
QY 8419 AAGGAAACTAAGAGGACACAAAGCTTTACTGTATCTGTAACACCCCTTATGATGAATCT 8478
Db 7732 AAGGAAACTAAGAGGACACAAAGCTTTACTGTATCTGTAACACCCCTTATGATGAATCT 7791
```

```
QY 8479 AAATTTTATATTTGGCTGTGATCGGTGTGCAGAAATTTGGTACCATGGCGCTGCGTTGGCATC 8538
Db 7792 AAATTTTATTTGGCTGTGATCGGTGTGCAGAAATTTGGTACCATGGCGCTGCGTTGGCATC 7851
QY 8539 TTGCAAGTGAAGCAGAGCTCAATGATGATGATGCTGTCCACAGTGCAGTGCAGTCAACAGAG 8598
Db 7852 TTGCAAGTGAAGCAGAGCTCAATGATGATGATGCTGTCCACAGTGCAGTGCAGTCAACAGAG 7911
QY 8599 GATGCATGACAGTGTCTCACGCCACTACAGCAAGAGGATTATGAGGGTTGAAGAGGGTG 8658
Db 7912 GATGCATGACAGTGTCTCACGCCACTACAGCAAGAGGATTATGAGGGTTGAAGAGGGTG 7971
QY 8659 CTCCTGTTCTTACAGGCCCATTAAGATGCGCTGGCTTCTTGTGAACCATAGACCTTAAT 8718
Db 7972 CTCCTGTTCTTACAGGCCCATTAAGATGCGCTGGCTTCTTGTGAACCATAGACCTTAAT 8031
QY 8719 GATGCACAGATTTATTTGTTGTTATTAAGGAACCTATGAGCTTGCCACCATGGAAGAA 8778
Db 8032 GATGCACAGATTTATTTGTTGTTATTAAGGAACCTATGAGCTTGCCACCATGGAAGAA 8091
QY 8779 AGAGTACAAAGACGATATTTATGAAAAGCTGACGGAATTTGTGCGAGATATGACCAAAAT 8838
Db 8092 AGAGTACAAAGACGATATTTATGAAAAGCTGACGGAATTTGTGCGAGATATGACCAAAAT 8151
QY 8839 TTTGATACTGTGTTTACTACAAATCAAGTACTCCCACTTTTACAGTGTGCAGAGTT 8898
Db 8152 TTTGATACTGTGTTTACTACAAATCAAGTACTCCCACTTTTACAGTGTGCAGAGTT 8211
QY 8899 CTCGAATCATTTTGTACAGAAATTTGAAGGCTTCAAGGCTAGCAGGCTCATACAAAC 8958
Db 8212 CTCGAATCATTTTGTACAGAAATTTGAAGGCTTCAAGGCTAGCAGGCTCATACAAAC 8271
QY 8959 AAACCTGCAGCTACAGCTTCTTAA 8982
Db 8272 AAACCTGCAGCTACAGCTTCTTAA 8295

RESULT 3
AK025776
LOCUS
DEFINITION
Homo sapiens cDNA: FLJ22123 fis, clone HEP19337, highly similar to
AB032251 Homo sapiens BPTF mRNA for bromodomain PHD finger
transcription factor.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3419)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (2992009) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency). Location/Qualifiers
FEATURES
```

[illegible]

QY	6255	TCCATCAAGTACAGGTACCAGT	CAGCAAAACCTTTACTTCA	TTCCAGCCGACGACGAAC	6314
Db	1920	TCCATCAAGTACAGGTACCAGT	CAGCAAAACCTTTACTTCA	TTCCAGCCGACGACGAAC	1379
QY	6315	AGTCACAAATTAGGCCCAATAC	CTCAGGCTCTGGAGGAACCA	CAGCAATTCACAAGTAAT	6374
Db	1980	AGTCACAAATTAGGCCCAATAC	CTCAGGCTCTGGAGGAACCA	CAGCAATTCACAAGTAAT	2039
QY	6375	CACAGGCGCTCAGATTCGCGC	CTGGTATGACCGGTATPAGA	CAACACCACTCCAACAGTCA	6434
Db	2040	CACAGGCGCTCAGATTCGCGC	CTGGTATGACCGGTATPAGA	CAACACCACTCCAACAGTCA	2099
QY	6435	ACTAGGAAGGCAATTAATTCGA	ACACCTGTGATGGTACACCG	GTATAGAACACCACTCCAAC	6494
Db	2100	ACTAGGAAGGCAATTAATTCGA	ACACCTGTGATGGTACACCG	GTATAGAACACCACTCCAAC	2159
QY	6495	GATGACTCAAAATCATCAGGG	GGCAGCGTCTCCACTGCACT	GTGAGGTACACCGAGGTCT	6554
Db	2160	GATGACTCAAAATCATCAGGG	GGCAGCGTCTCCACTGCACT	GTGAGGTACACCGAGGTCT	2219
QY	6555	TTCTCTAACACCTGGGGCAGAA	AGCTTAACTTCAGCAACGT	CTCACTTCAAAATATACAGT	6614
Db	2220	TTCTCTAACACCTGGGGCAGAA	AGCTTAACTTCAGCAACGT	CTCACTTCAAAATATACAGT	2279
QY	6615	TTGAGCTCTACAACCCCTCGCC	CCCAACAGGACAGTGAAGCT	CATCACCATGCGTCAACT	6674
Db	2280	TTGAGCTCTACAACCCCTCGCC	CCCAACAGGACAGTGAAGCT	CATCACCATGCGTCAACT	2339
QY	6675	TACTCAGTTAAACACAGGGCC	CAGCGTGCCAATCAAGGTTT	GACAGTAGTAGTAATTCAGG	6734
Db	2340	TACTCAGTTAAACACAGGGCC	CAGCGTGCCAATCAAGGTTT	GACAGTAGTAGTAATTCAGG	2399
QY	6735	AGGTCAAACTACTGGACAGTT	TTGAGTTGATACCTCAAGG	GGGTGACTGTACTCCAGGCCC	6794
Db	2400	AGGTCAAACTACTGGACAGTT	TTGAGTTGATACCTCAAGG	GGGTGACTGTACTCCAGGCCC	2459
QY	6795	AGCCAGCAGCTAATGCAAGCT	CAATGCCAAATGGTACTGTT	CAGCGATTCCTCTTTAC	6854
Db	2460	AGCCAGCAGCTAATGCAAGCT	CAATGCCAAATGGTACTGTT	CAGCGATTCCTCTTTAC	2519
QY	6855	CCCATTTGGCAACACAGSCCA	CCACAGCCAGCACCAACCAC	CACTGTTTCCACGACAGC	6914
Db	2520	CCCATTTGGCAACACAGSCCA	CCACAGCCAGCACCAACCAC	CACTGTTTCCACGACAGC	2579
QY	6915	AGCAGGTACAGTGAACAAAG	GAGTAGTAACTGTACCCCGA	TGATCATCAAGA	6974
Db	2580	AGCAGGTACAGTGAACAAAG	GAGTAGTAACTGTACCCCGA	TGATCATCAAGA	2639
QY	6975	CAAAACCTGCCACAGCTCAGT	CATCAAGTGTGGTCCAGCA	GAAGCCAGCCAGACAGC	7034
Db	2640	CAAAACCTGCCACAGCTCAGT	CATCAAGTGTGGTCCAGCA	GAAGCCAGCCAGACAGC	2699
QY	7035	TGCTCAGCCTTACAGCTCGGC	CCCAAGCCAGCCCGAGTCC	CCAGCTCAGCCTGA	7094
Db	2700	TGCTCAGCCTTACAGCTCGGC	CCCAAGCCAGCCCGAGTCC	CCAGCTCAGCCTGA	2759
QY	7095	AGTTCAGACTCAGCTGAAGTT	TCAGACCCAAACAAGTGT	TATATCCATGTCCTTCTGA	7154
Db	2760	AGTTCAGACTCAGCTGAAGTT	TCAGACCCAAACAAGTGT	TATATCCATGTCCTTCTGA	2819
QY	7155	AGCACACCCACCAACAGCAGT	CATCCAAAGCCCAAGTTG	CAGCAGTCTCAGCCTCA	7214
Db	2820	AGCACACCCACCAACAGCAGT	CATCCAAAGCCCAAGTTG	CAGCAGTCTCAGCCTCA	2879
QY	7215	AAGTAATGTCCAAGGACAGT	CTCCTGTTCGTGTCCTCA	AAAGTCCATCAGACTCGAAT	7274
Db	2880	AAGTAATGTCCAAGGACAGT	CTCCTGTTCGTGTCCTCA	AAAGTCCATCAGACTCGAAT	2939
QY	7275	TCCATCAACTCCATCCCACT	GTCTCCTGTGCAACAACAT	TCCAGGTTTCAGACTTCA	7334
Db	2940	TCCATCAACTCCATCCCACT	GTCTCCTGTGCAACAACAT	TCCAGGTTTCAGACTTCA	2999

QY	7335	ACACCCATTCCCAATTCACACCATACATACATCTCTTCAGATACCTTCCCAAGCCAGCCACA	7399
Db	3000	ACACCCATTCCCAATTCACACCATACATCTCTTCAGATACCTTCCCAAGCCAGCCACA	3059
QY	7395	GTCCACACCCCGAGGT	7409
Db	3060	GTCCACACCCCGAGGT	3074
RESULT 4			
HSU05237			
LOCUS	HSU05237	2673 bp	linear
DEFINITION	Human fetal Alz-50-reactive clone 1 (FAC1) mRNA, complete cds.		
ACCESSION	U05237		
VERSION	U05237.1	GI:1276427	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiinidae; Homo.		
AUTHORS	1 (bases 1 to 2673)		
TITLE	Bowser, R., Giambone, A. and Davies, P.		
JOURNAL	FAC1, a novel gene identified with the monoclonal antibody Alz50,		
MEDLINE	is developmentally regulated in human brain		
PUBMED	Dev. Neurosci. 17 (1), 20-37 (1995)		
REFERENCE	95347245		
AUTHORS	2 (bases 1 to 2673)		
TITLE	Bowser, R.P.		
JOURNAL	Direct Submission		
COMMENT	Submitted (18-JAN-1994) Robert P. Bowser, Dept. Pathology, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10641, USA		
FEATURES	On Apr 25, 1996 this sequence version replaced gi:451847.		
source	Location/Qualifiers		
gene	1..2673		
CDS	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="17"		
	/clone="fetal Alz-50-reactive clone 1"		
	/tissue_type="brain"		
	/clone_lib="Clontech HLI065b, fetal brain cDNA library in gtl1"		
	/dev_stage="gestational week 22"		
	1..2673		
	/gene="FAC1"		
	74..2506		
	/gene="FAC1"		
	/codon_start=1		
	/protein_id="AAA97522.1"		
	/db_xref="GI:1276428"		
	/translation="MVSEEEEDGDAEETQDSEDDDEDEDDDDSDYPEMEDDD		
	DDASYCTESSFRSHSYSTSPGRRKRVHRPRSPILKDKDIPLEPKSSEDLVMPVE		
	HIMNVAIYEVLRNFGTVLRSLRPFEDFCALVYSEQCTLMAEMHVLKAVLR		
	TSNTPTPADLKDSVNSTLYFIDGWTPEVLRVYCSDEKEYHVLVYQAEADYPYGP		
	ENKIKVLOFDVDFLTNIARELSMSEGVIOYDHCVRCHKGLDLCCTCSAVYHLE		
	CXKPIEVEDPESQCEVCAHVPGVTCVAEIOKNKPYIRHEPIGYDRSRKRYWFL		
	NRLIIEDTENEKKKIWYSTKVQLAELIDCLDKDYWEALCKLIEEMREIHRHM		
	DIETLNRKAGNSKFLAANEEI LESIRAKGDI DNKVSPEETEDKDNENTENSKS		
	AKNRREDFQSLKSDDKDTPDDPEQKSEPTVEVGDKNSVSNALGNTNTATSE		
	ETSPSGRSPVGLSETPDSNNMAKKVASELPQDVPPEPNTKTCESNTATTTQSP		
	NLENSNSSELSQSESAADDPENGERSHTPVSIQBEIVGDTSEKSTGSELSE		
	PGAGKSGSTRITITRLRNPDSKLSOLKSOQVAAAHAENKLPKECKEVLVNSQGEI		
	SRLSUTKKEVIMKGNINNYFKLQGEKRYRVYHNQYSNFSALNKHQHREHDKRRLAH		
	KFCUTPAGEFKFNQSVHSGSKVLITSLRTLTTOLETTSLHPSFIPTGHHIGQISGRQF		
	RCVANPENLHWL		
	2507..2653		
3'UTR	/gene="FAC1"		
BASE COUNT	887 a 530 c 678 g 578 t		
ORIGIN			
Query Match	26.0%	Score 2565.6;	DB 9; Length 2673;

[illegible]

Qy 2376 AGGGAAGATATCGGCTTACACCAATCAATACTCCACCAATTCATTTGCTTTTGAATAAGCA 2435
|||||
Db 2191 AGGGAAGATATCGGCTTACACCAATCAATACTCCACCAATTCATTTGCTTTTGAATAAGCA 2250
2436 CCAGCAGAGAGAGACCATATAGAGAGAGGCTCTTGGACATAAGTCTGTCTGACTCC 2495
|||||
Db 2251 CCAGCAGAGAGAGACCATATAGAGAGAGGCTCTTGGACATAAGTCTGTCTGACTCC 2310
2496 AGCAGGAGAGTCAAAATGGAGACGGTCTCTGTCATGGTCCAAAGTCTCTTACCATATCTAC 2555
|||||
Db 2311 AGCAGGAGAGTCAAAATGGAGACGGTCTCTGTCATGGTCCAAAGTCTCTTACCATATCTAC 2370
2556 TCTGAGACTGACTATCACCCTATAGAAAACATCCCTTCATCCTTTTTCATCCCAA 2615
|||||
Db 2371 TCTGAGACTGACTATCACCCTATAGAAAACATCCCTTCATCCTTTTTCATCCCAA 2429
2616 CTGGGCATCATAGGCAATTTGGATCAAGGAGTTCAGATGTGAGCAACCCAGAGA 2675
2430 CTGGGCATCATAGGCAATTTGGATCAAGGAGTTCAGATGTGAGCAACCCAGAGA 2489
2676 ATTTCATTTGGCTTTAGCCATTTGGAGTGTGAGTGTAAACCCAGTCTGTGATGCTACCAAT 2735
2490 ATTTCATTTGGCTTTAGCCATTTGGAGTGTGAGTGTAAACCCAGTCTGTGATGCTACCAAT 2549
2736 ATGCGGAGATTTTAGGACATCACCAGTTACCCGGATGACATCAATTTGAAAGAGAGA 2795
|||||
Db 2550 ATGCGGAGATTTTAGGACATCACCAGTTACCCGGATGACATCAATTTGAAAGAGAGA 2609
2796 AAAGGAGAAATGCAAAAAAAGAGAGAAACAGGAGAGAGAGAG 2839
2610 AAAGGAGAAATGCAAAAAAAGAGAGAAACAGGAGAGAGAGAG 2653

RESULT 5

AC107377/c
LOCUS AC107377 88848 bp DNA linear PRT-01-MAY-2002
DEFINITION Homo sapiens chromosome 17, clone CTD-2314M10, complete-sequence.
AC107377
VERSION AC107377.4 GI:20377031
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 88848)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone CTD-2314M10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 88848)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Marquis,N., Matthews,C.,
McClean,C., Macdonald,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,T., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

TITLE

JOURNAL
REFERENCE
AUTHORS

Submitted (19-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 88848)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Marquis,N., Matthews,C.,
McClean,C., Macdonald,P., McKernan,K., Meldrum,J., Meneus,L.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,T., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 88848)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Marquis,N., Matthews,C.,
McClean,C., Macdonald,P., McKernan,K., Meldrum,J., Meneus,L.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,T., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

TITLE
JOURNAL
COMMENT

Submitted (01-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 1, 2002 this sequence version replaced gi:20303884.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21459
Center clone name: 2314_M_10
----- Location/Qualifiers
1. 88848
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"

FEATURES
source

Location/Qualifiers
1. 88848
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"

repeat_region	/map="17"	16519..16524	unsure
repeat_region	/clone="CTD-2314M10"	/note="<30 Qual SNGL Region"	
repeat_region	/clone_lib="CITDI Human BAC"	16635..16933	
repeat_region	complement(98..394)	/rpt_family="AluJb"	
repeat_region	/rpt_family="AluSc"	complement(17254..17410)	
repeat_region	408..446	/rpt_family="MER58A"	
repeat_region	/rpt_family="L2"	17422..17547	
repeat_region	complement(1891..2157)	/rpt_family="L2"	
repeat_region	/rpt_family="AluSx"	complement(17944..18240)	
repeat_region	2253..2538	/rpt_family="AluSx"	
repeat_region	/rpt_family="L2"	complement(18426..18746)	
repeat_region	complement(2618..2816)	/rpt_family="AluSx"	
repeat_region	/rpt_family="AluSx"	complement(18751..19065)	
repeat_region	2882..3044	/rpt_family="AluSx"	
repeat_region	/rpt_family="L2"	19095..19142	
repeat_region	3053..3253	/rpt_family="AT_rich"	
repeat_region	/rpt_family="MIR"	complement(19477..19574)	
repeat_region	complement(4123..4358)		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(4539..4848)		
repeat_region	/rpt_family="AluJb"		
repeat_region	complement(5101..5396)		
repeat_region	/rpt_family="AluSc"		
repeat_region	complement(5449..5754)		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(5927..6211)		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(6398..6702)		
repeat_region	/rpt_family="AluJo"		
repeat_region	complement(6715..7020)		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(7275..7572)		
repeat_region	/rpt_family="AluSx"		
repeat_region	7725..7745		
repeat_region	/rpt_family="AT_rich"		
repeat_region	8488..8513		
repeat_region	/rpt_family="(TTTA)n"		
repeat_region	complement(8517..8666)		
repeat_region	/rpt_family="FRAM"		
repeat_region	8667..8695		
repeat_region	/rpt_family="(TTCA)n"		
repeat_region	complement(8714..9005)		
repeat_region	/rpt_family="AluSc"		
repeat_region	complement(9022..9328)		
repeat_region	/rpt_family="AluSx"		
repeat_region	9492..9785		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(9849..10156)		
repeat_region	/rpt_family="AluJb"		
repeat_region	11807..12100		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(12718..12832)		
repeat_region	/rpt_family="MIR"		
repeat_region	13094..13244		
repeat_region	/rpt_family="FRAM"		
repeat_region	13367..13673		
repeat_region	/rpt_family="AluJ"		
repeat_region	13686..13997		
repeat_region	/rpt_family="AluJo"		
repeat_region	14455..14475		
repeat_region	/rpt_family="AT_rich"		
repeat_region	complement(14478..14608)		
repeat_region	/rpt_family="FLAM_C"		
repeat_region	14917..14938		
repeat_region	/rpt_family="(A)n"		
repeat_region	14946..15215		
repeat_region	/rpt_family="AluJb"		
repeat_region	15576..15735		
repeat_region	/rpt_family="AluSx"		
repeat_region	15736..15982		
repeat_region	/rpt_family="AluSx"		
repeat_region	16211..16518		
repeat_region	/rpt_family="AluSx"		
<hr/>			
Query Match 23.5%; Score 2316; DB 9; Length 88848;			
Best Local Similarity 99.6%; Pred. No. 0;			
Matches 2322; Conservative 0; Mismatches 10; Indels 0; Gaps 0;			
Qy	3210	AGATGTGAAGGAGCTCTAGATTCTGACAGTGATAAACCTGCAAGGAAGAACCAATGCA 3269	
Db	29115	AGATGTGAAGGAGCTCTTAGATTCTGACAGTGATAAACCTGCAAGGAAGAACCAATGCA 29056	
Qy	3270	AGTAGACGATGACATGAAACACAGAGTCACATGTAATGTGAGGAGAGTTCTCAAGTAGA 3329	
Db	29055	AGTAGACGATGACATGAAACACAGAGTCACATGTAATGTGAGGAGAGTTCTCAAGTAGA 28996	
Qy	3330	TGTGGTCAATGTTAGTGAGGGTTTTCATCTAAGGACTAGTTACAAAAGAAAACAAAATC 3389	
Db	28995	TGTGGTCAATGTTAGTGAGGGTTTTCATCTAAGGACTAGTTACAAAAGAAAACAAAATC 28936	
Qy	3390	ATCCAAACTAGATGGACTTCTGAAAGGAGAAATTAACACAGTTTACACTGGAAGAAAACA 3449	
Db	28935	ATCCAAACTAGATGGACTTCTGAAAGGAGAAATTAACACAGTTTACACTGGAAGAAAACA 28876	
Qy	3450	CGGACTCGAAAAATCAAGTTGGAGGTGGAATTAAGGGTATAGGAAGAACTTCTACAA 3509	
Db	28875	CGGACTCGAAAAATCAAGTTGGAGGTGGAATTAAGGGTATAGGAAGAACTTCTACAA 28816	
Qy	3510	TTCTTCAAAAATCTCTGAAATCACCAGTAATAACGAAAGAAAAGAGGGTGTTCAGAG 3569	
Db	28815	TTCTTCAAAAATCTCTGAAATCACCAGTAATAACGAAAGAAAAGAGGGTGTTCAGAG 28756	
Qy	3570	TGACTCGATGAGACAAGAACAGACCCCAATGCAATTAATGATCAACCTGAGACTTGAT 3629	
Db	28755	TGACTCGATGAGACAAGAACAGACCCCAATGCAATTAATGATCAACCTGAGACTTGAT 28696	
Qy	3630	TCAGGAGTTTCACAAAGTATTCCTCAGTTCTTAGAATGAGTGATCCTAGTATACACAC 3689	
Db	28695	TCAGGAGTTTCACAAAGTATTCCTCAGTTCTTAGAATGAGTGATCCTAGTATACACAC 28636	
Qy	3690	AAACAAACTTTATCCAAAGATCGAGTGTGATGATGTCTCCATTCGGAGCCCAAGAAC 3749	
Db	28635	AAACAAACTTTATCCAAAGATCGAGTGTGATGATGTCTCCATTCGGAGCCCAAGAAC 28576	
Qy	3750	AAAATGTCGGAACAAAATTCATTTGAAATGACATAGAGAAAAGTCTGTGACCTTGC 3809	
Db	28575	AAAATGTCGGAACAAAATTCATTTGAAATGACATAGAGAAAAGTCTGTGACCTTGC 28516	
Qy	3810	CAGTAGAGCCAGGAACCCACTTAAGAGTAAAAACCAAGAAATGATTTTTTCATCGATGA 3869	
Db	28515	CAGTAGAGCCAGGAACCCACTTAAGAGTAAAAACCAAGAAATGATTTTTTCATCGATGA 28456	
Qy	3870	CTCTAAACTAGCCAGTGCAGATGATATTGGTACTTGTATGATCTGTAGAACAAAACCCGT 3929	
Db	28455	CTCTAAACTAGCCAGTGCAGATGATATTGGTACTTGTATGATCTGTAGAACAAAACCCGT 28396	
Qy	3930	CATACAGGAGGAAGTGACACCATTTCTTCTTCCAGAGTGTCTTACATTCATCAGT 3989	
Db	28395	CATACAGGAGGAAGTGACACCATTTCTTCTTCCAGAGTGTCTTACATTCATCAGT 28336	

Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,
 Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T.,
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (42-OCT-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 108860)
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,
 Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T.,
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (30-OCT-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 30, 2002 this sequence version replaced gi:23915491.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RN/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L27577
 Center clone name: 551_H_18

 Location/Qualifiers
 1. .108860
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /Clone="RP13-551H18"
 /Clone_lib="RPC1-13 Human Female BAC"
 complement(1. .61)
 /rpt_family="AluY"
 81. 240
 /rpt_family="MIR"
 241. 537
 /rpt_family="AluSx"
 569. 592
 /rpt_family="AT-rich"
 607. 898
 /rpt_family="AluYb"
 complement(904. .950)
 /rpt_family="MLT1F1"
 complement(951. .1230)
 /rpt_family="AluSg"
 complement(1231. .1569)
 /rpt_family="MLT1F1"
 1570. .1685
 /rpt_family="L1Pb1"
 1686. .1713
 /rpt_family="(TG)n"
 1713. .1743
 /rpt_family="(TA)n"
 1714. .2101
 /rpt_family="L1Pb1"
 complement(2102. .2277)
 /rpt_family="MLT1F1"
 complement(2292. .2576)
 /rpt_family="AluSg"
 2707. .2732
 /rpt_family="AT-rich"
 complement(2795. .3089)
 /rpt_family="AluSx"
 3092. .3121
 /rpt_family="(TA)n"
 complement(3122. .3419)
 /rpt_family="AluSx"
 3448. .3758
 /rpt_family="AluSc"
 3869. .4006
 /rpt_family="FLAM_C"
 4153. .4489
 /rpt_family="AluYb"
 4628. .4932
 /rpt_family="AluY"
 5080. .5390
 /rpt_family="AluY"
 5407. .5701
 /rpt_family="AluSp"
 5743. .6051
 /rpt_family="AluSg"
 6058. .6362
 /rpt_family="AluY"
 6366. .6667
 /rpt_family="AluY"
 6672. .6818
 /rpt_family="L1M3"
 6822. .7143
 /rpt_family="AluSx"
 7238. .7338
 /rpt_family="L1M3"
 7342. .7433
 /rpt_family="(TA)n"
 7460. .7749
 /rpt_family="AluSg"
 7851. .7970
 /rpt_family="L1M3"
 complement(7971. .8272)
 /rpt_family="AluSx"
 8273. .8322
 /rpt_family="L1M3"
 8331. .8394
 /rpt_family="MER21B"
 complement(8412. .8563)
 /rpt_family="AluYb"
 8587. .8618
 /rpt_family="L2"
 complement(8980. .9289)
 /rpt_family="AluSg"
 complement(9866. .10174)
 /rpt_family="AluSg"
 complement(10220. .10548)
 /rpt_family="AluY"
 complement(10594. .10893)

[illegible]

Db	95496	AGAAAATTGTGCAAAATCCCACTGTCAACACCACCTACAAAGTACCAAGCTTTCAC	95437	
Qy	5130	ACCCTCACAGGGCGGAGTGTGGACATCATCTCTGTAAAGGACGACGCAAAACCGTGGT	5189	
Db	95436	ACCCTCACAGGGCGGAGTGTGGACATCATCTCTGTAAAGGACGACGCAAAACCGTGGT	95377	
Qy	5190	CACCACACAGTGACACGACCTCCCTGACCACCGGAGGACACACTGGTTACATCTATGAC	5249	
Db	95376	CACCACACAGTGACACGACCTCCCTGACCACCGGAGGACACACTGGTTACATCTATGAC	95317	
Qy	5250	TGTGAGCAAAAGTATTCCACGACGACAAAGTGAACACTGATGAAATTTTCAGACACAAA	5309	
Db	95316	TGTGAGCAAAAGTATTCCACGACGACAAAGTGAACACTGATGAAATTTTCAGACACAAA	95257	
Qy	5310	GAACACTGCTTCAGGTACAGTCTGCCATCTATAGAAAATTTGTTACCAAGACGACCAA	5369	
Db	95256	GAACACTGCTTCAGGTACAGTCTGCCATCTATAGAAAATTTGTTACCAAGACGACCAA	95197	
Qy	5370	GAAGAGCATTTTCTTTTGGCTAATGATGACTTAAAAAAGTTGGCCCGAAAAGGAGGAAT	5429	
Db	95196	GAAGAGCATTTTCTTTTGGCTAATGATGACTTAAAAAAGTTGGCCCGAAAAGGAGGAAT	95137	
Qy	5430	CCGAGAGTCCCTTATTTTAAATACAAATGCAAAACCTGCTTGGATATATGGCCATATCC	5489	
Db	95136	CCGAGAGTCCCTTATTTTAAATACAAATGCAAAACCTGCTTGGATATATGGCCATATCC	95077	
Qy	5490	TTCTCCTAGACCGACCTTTGGCATCACTTGGAGGTATAGACTTCAGACAGTA	5541	
Db	95076	TTCTCCTAGACCGACCTTTGGCATCACTTGGAGGTATAGACTTCAGACAGTA	95025	
RESULT 7				
AC134407		199875 bp	DNA	linear
LOCUS	AC134407			PRI 07-JUN-2003
DEFINITION	AC134407			Homo sapiens chromosome 17, clone Rpl1-855A2, complete sequence.
ACCESSION	AC134407.6		Gr:31455710	
VERSION				
KEYWORDS	HTG.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
AUTHORS				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
AUTHORS				
REFERENCE				
AUTHORS				

Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Monaghan, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 199875)

REFERENCE
AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Monaghan, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

5 (bases 1 to 199875)

REFERENCE
AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Monaghan, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

6 (bases 1 to 199875)

REFERENCE
AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,

Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melrim, J., Meneus, L., Miho, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Rhomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (08-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
7 (bases 1 to 199875)

TITLE JOURNAL REFERENCE AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melrim, J., Meneus, L., Miho, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Rhomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (07-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 7, 2003 this sequence version replaced gi:29611683.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L28171
Center clone name: 855_A_2

FEATURES source

Location/Qualifiers
1. .199875
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone_lib="RP11-855A2"
/clone="RP11-855A2"
445..494
repeat_region
/rpt_family="AT-rich"
893..1215
repeat_region
/rpt_family="AluY"
1434..1465
repeat_region
/rpt_family="(T)n"
complement(1551..1631)
repeat_region
/rpt_family="MIR"

complement(1952..2262)
/rpt_family="AluYa5"
2435..2654
/rpt_family="AluS9/x"
complement(3014..3317)
/rpt_family="AluSx"
3624..3926
/rpt_family="AluSx"
4010..4291
/rpt_family="AluJo"
4368..4677
/rpt_family="AluSq"
4695..5184
/rpt_family="L1MD3"
repeat_region
5185..5233

Query Match 23.5%; Score 2314.4; DB 9; Length 199875;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2321; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 3210 AGATGTGAAGGAGCTCTTAGATTCTGACAGTGTATAAACCTGCAAGGAAGAACAATGGA 3269
DB 42003 AGATGTGAAGGAGCTCTTAGATTCTGACAGTGTATAAACCTGCAAGGAAGAACAATGGA 42062
QY 3270 AGTAGACGATGACATGAAAAACAGAGTCACATGTAAATTGTCAGGAGAGTTCTCAAGTAGA 3329
DB 42063 AGTAGACGATGACATGAAAAACAGAGTCACATGTAAATTGTCAGGAGAGTTCTCAAGTAGA 42122
QY 3330 TGTGTCATGTTAGTGAGGGTTTTCATCTAAGGACTAGTTACAAAAAGAAACAANAATC 3389
DB 42123 TGTGTCATGTTAGTGAGGGTTTTCATCTAAGGACTAGTTACAAAAAGAAACAANAATC 42182
QY 3390 ATCCAAACTAGATGGACTTCTTGAAGAGAGAATTAACACTGTATACACTGGAAGAAAAACA 3449
DB 42183 ATCCAAACTAGATGGACTTCTTGAAGAGAGAATTAACACTGTATACACTGGAAGAAAAACA 42242
QY 3450 GGCAGTCGAAAAAATCAAGTTGGAGGGTGGAAATTAAGGGTATAGGAAAGACTTCTCAAA 3509
DB 42243 GGCAGTCGAAAAAATCAAGTTGGAGGGTGGAAATTAAGGGTATAGGAAAGACTTCTCAAA 42302
QY 3510 TTCTTCAAAAATCTCTGAATCACCAGTAAACAGAAAGCAAGAAAGAGGGTGTCTAGAG 3569
DB 42303 TTCTTCAAAAATCTCTGAATCACCAGTAAACAGAAAGCAAGAAAGAGGGTGTCTAGAG 42362
QY 3570 TGACTCGATGAGCAAGAAGACAGAGCCCAATGCAATTAATGATCAACCTGAGGACTTGTAT 3629
DB 42363 TGACTCGATGAGCAAGAAGACAGAGCCCAATGCAATTAATGATCAACCTGAGGACTTGTAT 42422
QY 3630 TCAGGAGTCTTCAAAAAGTGTCTTCAGTTCTTTAGAAATGAGTGTAGTCTTAGTCAATACCAC 3689
DB 42423 TCAGGAGTCTTCAAAAAGTGTCTTCAGTTCTTTAGAAATGAGTGTAGTCTTAGTCAATACCAC 42482
QY 3690 AAACAACCTTTATCCAAAAGATCGAGTGTAGATGATGTCTCCATTCGAGGCCCAAAAC 3749
DB 42483 AAACAACCTTTATCCAAAAGATCGAGTGTAGATGATGTCTCCATTCGAGGCCCAAAAC 42542
QY 3750 AAAATGTCGAAAAACAAAATTCATTGAAAATGACATAGAGAAGAAAAGTCTCTGACCTTGC 3809
DB 42543 AAAATGTCGAAAAACAAAATTCATTGAAAATGACATAGAGAAGAAAAGTCTCTGACCTTGC 42602
QY 3810 CAGTAGAGCCAGGAACCCCACTAAGAGTAAAACCAAGGAATGATTTTTTTCATCGATGA 3869
DB 42603 CAGTAGAGCCAGGAACCCCACTAAGAGTAAAACCAAGGAATGATTTTTTTCATCGATGA 42662
QY 3870 CTCCTAACTAGCCAGTGCAGATCATATTGTTGATCTCTGATCTGTAAGAAACAAAACCCGCT 3929
DB 42663 CTCCTAACTAGCCAGTGCAGATCATATTGTTGATCTCTGATCTGTAAGAAACAAAACCCGCT 42722
QY 3930 CATACAGGAGGAAGTGACACCATTTCTTCTTCTTCCAAAGAGTGTCTTTTACATTCATCAGT 3989
DB 42723 CATACAGGAGGAAGTGACACCATTTCTTCTTCTTCCAAAGAGTGTCTTTTACATTCATCAGT 42782
QY 3990 GCCTAAAGTACCAATGACAGAGATGCCACACCTCTGTCAAGAGCAATGGACTTTTGAAG 4049

QY	6777	GACTGTACTCCAGGCCAGCCAGCAGCTAAATGCAAGCTGCAATGCCAAATGGTACTGT	6836
DB	13	GACTGTACTCCAGGCCAGCCAGCAGCTAAATGCAAGCTGCAATGCCAAATGGTACTGT	72
QY	6837	TCAGCGAATCCTCTTTACCCCATTTGGCAACAACAGCCACACAGCCAGCCAGCCACC	6896
DB	73	TCAGCGAATCCTCTTTACCCCATTTGGCAACAACAGCCACACAGCCAGCCAGCCACC	132
QY	6897	CACGTGTTTCCACGACAGCAGCAGGTACAGGTGAACAAAGGACAGAGTAAACTGTACACCCCA	6956
DB	133	CACGTGTTTCCACGACAGCAGCAGGTACAGGTGAACAAAGGACAGAGTAAACTGTACACCCCA	192
QY	6957	GATCGAGGTACATCAAGACAAAACCCCTGCCACAGCTCAGTCATCAAGTGTGGTCCAGC	7016
DB	193	GATCGAGGTACATCAAGACAAAACCCCTGCCACAGCTCAGTCATCAAGTGTGGTCCAGC	252
QY	7017	AAAAGCCAGCCACAGACTGCTCAGCCTTCAGCTCGCCGCCAGCCCAACCCAGCCCA	7076
DB	253	AGAAGCCAGCCACAGACTGCTCAGCCTTCAGCTCGCCGCCAGCCCAACCCAGCCCA	312
QY	7077	GTCCCCAGCTCAGCCTGAAGTTTCAGACTCAGCCTGAAGTTTCAGACCCCAACAACTGTTTC	7136
DB	313	GTCCCCAGCTCAGCCTGAAGTTTCAGACTCAGCCTGAAGTTTCAGACCCCAACAACTGTTTC	372
QY	7137	ATCCATGTCCCTTCTGAAGACAACCCAGCCAGCAGCTCATATCCAAAGCCCAAGTTGC	7196
DB	373	ATCCATGTCCCTTCTGAAGACAACCCAGCCAGCAGCTCATATCCAAAGCCCAAGTTGC	432
QY	7197	AGCACAGCTCAGCCTCAAAAGTAAATGTCGAAGGACAGCTCTCTGCTGTCGAAGTCC	7256
DB	433	AGCACAGCTCAGCCTCAAAAGTAAATGTCGAAGGACAGCTCTCTGCTGTCGAAGTCC	492
QY	7257	ATCACAGACTCGAATAGCTCCATCAACTCCATCCCAACTGTCCTGGACAAACAAATCCCA	7316
DB	493	ATCACAGACTCGAATAGCTCCATCAACTCCATCCCAACTGTCCTGGACAAACAAATCCCA	552
QY	7317	GGTTCAGACTACAACCTCACAACCGATTCGAATTCGAACACATACATCTCTTCAGATACC	7376
DB	553	GGTTCAGACTACAACCTCACAACCGATTCGAATTCGAACACATACATCTCTTCAGATACC	612
QY	7377	TTCCCAAGGCGCAGCAGTCACAACCCAGGTACAGTCTTCAACTCAAACTCTTTCATC	7436
DB	613	TTCCCAAGGCGCAGCAGTCACAACCCAGGTACAGTCTTCAACTCAAACTCTTTCATC	640
QY	7437	AGGACAAACTTTAAATCAAGTTAGTGTTCATCCCATCCCTCCCTCAGCTACAAAATACA	7496
DB	641	AGGACAAACTTTAAATCAAGTTAGTGTTCATCCCATCCCTCCCTCAGCTACAAAATACA	640
QY	7497	GCAGCCACAGCCCAAGTCAATGGTGTGCCTCAGCTGCAACACAAGTCCAGGTTCTCTC	7556
DB	641	GCAGCCACAGCCCAAGTCAATGGTGTGCCTCAGCTGCAACACAAGTCCAGGTTCTCTC	640
QY	7557	TCAGATCCAGTCACAGGTTGTGCTCAGATACAGGCTCACAAGTGGTGTGCCCCAGCA	7616
DB	641	TCAGATCCAGTCACAGGTTGTGCTCAGATACAGGCTCACAAGTGGTGTGCCCCAGCA	640
QY	7617	AATCAAACTCCAGTTACCTATCCAAATTCAGAAAGCAGTGCCTGCAGACTCACCAGAT	7676
DB	641	AATCAAACTCCAGTTACCTATCCAAATTCAGAAAGCAGTGCCTGCAGACTCACCAGAT	640
QY	7677	TCAGAAATGTGGTTACAGTCGAGGAGCCAGTGTGCAAGAGCAGTTCGAAAGGTTACAGCA	7736
DB	641	TCAGAAATGTGGTTACAGTCGAGGAGCCAGTGTGCAAGAGCAGTTCGAAAGGTTACAGCA	640
QY	7737	ACTCAGGATCAGCAGCAAAAGAAAGAACAGACAGATAGAAATTAACGTGAACACACC	7796
DB	641	ACTCAGGATCAGCAGCAAAAGAAAGAACAGACAGATAGAAATTAACGTGAACACACC	640
QY	7797	CTCCAAGCTTCTAATCAAAAGTTGAAATCATTCAGAAACAGGTGGTGTGAAGCATATGC	7856
DB	641	CTCCAAGCTTCTAATCAAAAGTTGAAATCATTCAGAAACAGGTGGTGTGAAGCATATGC	663

Db	1744		TGCCACCATGGAGAGATACAAAGACGATATTATGAAAAGCTGACGGAATTTGTGC	1803
Qy	8823		AGATATGACCAAAATTTTGTAACTGCTGTTACTACAAATCAAGTGAAGTCCCAATTTTA	8882
Db	1804		AGATATGACCAAAATTTTGTAACTGCTGTTACTACAAATCAAGTGAAGTCCCAATTTTA	1863
Qy	8883		CCAGTGTGAGAGTCTCGAATCATCTTTGTACAGAAATTCARAGGCTTCAAGCTAG	8942
Db	1864		CCAGTGTGAGAGTCTCGAATCATCTTTGTACAGAAATTCARAGGCTTCAAGCTAG	1923
Qy	8943		CAGTCTCATACAAACAACTGCACTACAGCTCTTAAAGTTCAGCGTGTAACTAA	9002
Db	1924		CAGTCTCATACAAACAACTGCACTACAGCTCTTAAAGTTCAGCGTGTAACTAA	1983
Qy	9003		CATAAAACACAGCAAGAACTGTTGCTGCACTATTTAAATTAAGAGCCAGATGTTT	9062
Db	1984		CATAAAACACAGCAAGAACTGTTGCTGCACTATTTAAATTAAGAGCCAGATGTTT	2043
Qy	9063		TTAGTCAGGCTATCCTGACAGACTTGACCTAACTTCGTTTTATTGGTCAACAGTC	9122
Db	2044		TTAGTCAGGCTATCCTGACAGACTTGACCTAACTTCGTTTTATTGGTCAACAGTC	2103
Qy	9123		CAATTAATATCTTGGCCAAATTTTGTCCACGGACAAAGAAAAGTCAACGACACC	9182
Db	2104		CAATTAATATCTTGGCCAAATTTTGTCCACGGACAAAGAAAAGTCAACGACACC	2163
Qy	9183		ATTATCTTGCAGATCAGATGTTTGTGCTGCACTATTTGTCGACAGCGAGAACTTTCTTA	9242
Db	2164		ATTATCTTGCAGATCAGATGTTTGTGCTGCACTATTTGTCGACAGCGAGAACTTTCTTA	2223
Qy	9243		TTGAAAAAAGAAAAAGAAAGCAAGAAAAAGATACCTATGGGTCAGTGTAACTCC	9302
Db	2224		TTGAAAAAAGAAAAAGAAAGCAAGAAAAAGATACCTATGGGTCAGTGTAACTCC	2283
Qy	9303		ATGGAATGGCACTGCTCTTCTAGTGAAGAGCTGGTTTAGAGTCTCACAAAACTT	9362
Db	2284		ATGGAATGGCACTGCTCTTCTAGTGAAGAGCTGGTTTAGAGTCTCACAAAACTT	2343
Qy	9363		TTGACTGATTTATTTATTTGTCGAAAAAGAGCGCTTTTATTGCTGCGCCTCATTTGTC	9422
Db	2344		TTGACTGATTTATTTATTTGTCGAAAAAGAGCGCTTTTATTGCTGCGCCTCATTTGTC	2403
Qy	9423		AGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTTACATATAATCATCTGATCTTAT	9482
Db	2404		AGCTAAGTATTTTCTTATAAAATCCAGCCCGGTTTACATATAATCATCTGATCTTAT	2463
Qy	9483		CATGATTCCTGTAGTAAAGTACAGACGACCTCTAGATGCTCTTTCTTCTATGAAG	9542
Db	2464		CATGATTCCTGTAGTAAAGTACAGACGACCTCTAGATGCTCTTTCTTCTATGAAG	2523
Qy	9543		GAGTGTCTATCTACACATGTGCACACACACAACTGGGAATCAACAAATGAGTTTATGT	9602
Db	2524		GAGTGTCTATCTACACATGTGCACACACACAACTGGGAATCAACAAATGAGTTTATGT	2583
Qy	9603		TCATGGTATATAAAATTAAGTGTGCTAAGTGGCTAAGTGGCTTGGGCTACAG	9662
Db	2584		TCATGGTATATAAAATTAAGTGTGCTAAGTGGCTAAGTGGCTTGGGCTACAG	2642
Qy	9663		ACTCTGTTGCGCTGGAATATACAGTACAAATTTGTCAATTAATCTCTGCACGAGCTAAAGTG	9722
Db	2643		ACTCTGTTGCGCTGGAATATACAGTACAAATTTGTCAATTAATCTCTGCACGAGCTAAAGTG	2702
Qy	9723		AGTAAATCTATTTGAAGGTATCTTGTGTTGTAACATTTGTCAGATTTCTAATTTTTTCT	9782
Db	2703		AGTAAATCTATTTGAAGGTATCTTGTGTTGTAACATTTGTCAGATTTCTAATTTTTTCT	2762
Qy	9783		TTTGTATTAATTAATCACTATGGATGTATATGAACAAAAATTAATGGAGATAATTTTCT	9842
Db	2763		TTTGTATTAATTAATCACTATGGATGTATATGAACAAAAATTAATGGAGATAATTTTCT	2822
Qy	9843		CCACAAAAAA 9854	

Db 2823 CCCACAGACAGA 2834

RESULT 9
BD157797
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BD157797
Primer for synthesizing full-length cDNA and use thereof.
BD157797
BD157797.1 GI:27863555
JP 2002191363-A/12640.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 12640 09-JUL-2002
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/12640
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(387). .(2057).
FT CDS Location/Qualifiers
source
1. .2057
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 666 a 566 c 435 g 390 t
ORIGIN
Query Match 17.4%; Score 1719.4; DB 6; Length 2057;
Best Local Similarity 99.7%; Pred. No. 1.8e-265;
Matches 1723; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 5681 TAATTAAGAGGAGAGATGTTGGTCTTATGGCAATTCGATTTGATTAATTTGATCAGGAAA 5740
Db 1 TAATTAAGAGGAGAGATGTTGGTCTTATGGCAATTCGATTTGATTAATTTGATCAGGAAA 60
Qy 5741 TCATTTGTCCCATTTGGAGTTCCAGAACACCAAGAAAGCGCTACACCTCAGAGAAAG 5800
Db 61 TCATTTGTCCCATTTGGAGTTCCAGAACACCAAGAAAGCGCTACACCTCAGAGAAAG 120
Qy 5801 GCCTTCGATCAAGTGCAGTCCGCGCCAAAGAGACAGAAAGCGCCCAAGCAAACTGGCCCTG 5860
Db 121 GCCTTCGATCAAGTGCAGTCCGCGCCAAAGAGACAGAAAGCGCCCAAGCAAACTGGCCCTG 180
Qy 5861 TTATTAATTAAGAACTGGGTAGCAGAAAGAACTGGAATTTGGGAGATCAGGCAATTTG 5920
Db 181 TTATTAATTAAGAACTGGGTAGCAGAAAGAACTGGAATTTGGGAGATCAGGCAATTTG 240
Qy 5921 CTGAGAGTGGAGAAAAGGACAAAGCAAGCAAGTTGAGCAACAGCGCTTAAGAAACGACTGG 5980
Db 241 CTGAGAGTGGAGAAAAGGACAAAGCAAGCAAGTTGAGCAACAGCGCTTAAGAAACGACTGG 300
Qy 5981 AGCAGAGAAAGCCGACAGTGAATTCGCACTTCCACTACTTCCCAACAAAGCAGTACACCA 6040
Db 301 AGCAGAGAAAGCCGACAGTGAATTCGCACTTCCACTACTTCCCAACAAAGCAGTACACCA 360
Qy 6041 GCACCATCTCTCCAGCAGCAAGAAAGTTATGGTGGCCCAATTAAGTGGCTCAGTTTACAACTG 6100
|||||

Db 361 GCACCATCTCTCCAGCACAGAGGTTATGGTGGCCCCCATTAAGTGGCTCAGTTCAACATG 420
QY 6101 GAACCAAAATGTTACTAATAAGTTGGATCTCCAGCTACAGTAACATTCACACAA 6160
Db 421 GAACCAAAATGTTACTAATAAGTTGGATCTCCAGCTACAGTAACATTCACACAA 480
QY 6161 ACAAGAACCTTTCATCAAACTTTGTTACATGGTTAAAGCAAGCCAGTCATAATTCAGGGC 6220
Db 481 ACAAGAGCTTTTCATCAAACTTTGTTACATGGTTAAAGCAAGCCAGTCATAATTCAGGGC 540
QY 6221 TTGTTCAAGTACAGCAGAAAGTCCTGGGTATCATTCATCAAGTACAGTACAGTCAGC 6280
Db 541 TTGTTCAAGTACAGCAGAAAGTCCTGGGTATCATTCATCAAGTACAGTACAGTCAGC 600
QY 6281 AAACCTTTTACTTTCAGCCAGGACAGCAGCAACAGTACAAATAGGCCCAATACCTCAG 6340
Db 601 AAACCTTTTACTTTCAGCCAGGACAGCAGCAACAGTACAAATAGGCCCAATACCTCAG 660
QY 6341 GCTCTGGAGAACCAACCAAGCAATTCACAAGTAAATCACAGGGCTCAGATTCGCCCTGGTA 6400
Db 661 GCTCTGGAGAACCAACCAAGCAATTCACAAGTAAATCACAGGGCTCAGATTCGCCCTGGTA 720
QY 6401 TGACCGTGTATTAGAACACACCACTCCAAACAGTCAACACTAGGAAAGCAATTAATTCGAACAC 6460
Db 721 TGACCGTGTATTAGAACACACCACTCCAAACAGTCAACACTAGGAAAGCAATTAATTCGAACAC 780
QY 6461 CTGTGATGTTACAGCCAGGTCCTCTCAGCAAGTATGATCAAAATCATCAGGGGGCAGC 6520
Db 781 CTGTGATGTTACAGCCAGGTCCTCTCAGCAAGTATGATCAAAATCATCAGGGGGCAGC 840
QY 6521 CTGTCTCACTCAGTCTCCGCCCTTAACACAGGTTTCTCAACACTTGGGCAAGAAAGCT 6580
Db 841 CTGTCTCACTCAGTCTCCGCCCTTAACACAGGTTTCTCAACACTTGGGCAAGAAAGCT 900
QY 6581 TAACTTTCAGCAAGTCCACTTCAAAATATACAGTCTTACGCCCTCAACACTTGGGCAAGAAAGCT 6640
Db 901 TAACTTTCAGCAAGTCCCGTTCAAAATATACAGTCTTACGCCCTCAACACTTGGGCAAGAAAGCT 960
QY 6641 AACAGGACAAGTGAAGCTCACCATTGCTTCACTTACTCAGTTAAACACAGGGCCACGGTG 6700
Db 961 AACAGGACAAGTGAAGCTCACCATTGCTTCACTTACTCAGTTAAACACAGGGCCACGGTG 1020
QY 6701 GCAATCAAGGTTTGACAGTAGTAATTCAGGACACAGGTCAAACTACTGGACAGTTGCAAT 6760
Db 1021 GCAATCAAGGTTTGACAGTAGTAATTCAGGACACAGGTCAAACTACTGGACAGTTGCAAT 1080
QY 6761 TGATACCTCAAGGGGTGACTGTACCCAGGCCCCAGGCAGCAGCTAATGCAAGCTGCAA 6820
Db 1081 TGATACCTCAAGGGGTGACTGTACCCAGGCCCCAGGCAGCAGCTAATGCAAGCTGCAA 1140
QY 6821 TGCCAAATGGTACTGTTCAGCGATTCTCTTTACCCCTATTGGCAACACAGCCACACAG 6880
Db 1141 TGCCAAATGGTACTGTTCAGCGATTCTCTTTACCCCTATTGGCAACACAGCCACACAG 1200
QY 6881 CCAGCACCACACACCACTGTTTCCAGCAGCAGCAGGTACAGGTGAACAAAGGCAGA 6940
Db 1201 CCAGCACCACACCACTGTTTCCAGCAGCAGCAGGTACAGGTGAACAAAGGCAGA 1260
QY 6941 GTAATCTGTCACCCAGATGACAGGTACATCAAGACAAACCTGCCACCGCTCAGTCAT 7000
Db 1261 GTAATCTGTCACCCAGATGACAGGTACATCAAGACAAACCTGCCACCGCTCAGTCAT 1320
QY 7001 CAAAGTGTGGTCCAGCAAAAGCCAGCCAGCAGCTGCTCAGCTTTCAGCTCGGCCCCAGC 7060
Db 1321 CAAAGTGTGGTCCAGCAAAAGCCAGCCAGCAGCTGCTCAGCTTTCAGCTCGGCCCCAGC 1380
QY 7061 CCCAAACCCAGCCAGTCCCGAGCTCAGCTGAAAGTTTCAAGTCTCAGCTCAGCTGAAAGTTCAGA 7120
Db 1381 CCCAAACCCAGCCAGTCCCGAGCTCAGCTGAAAGTTTCAAGTCTCAGCTCAGCTGAAAGTTCAGA 1440
QY 7121 CCCAAACCACTGTTTCATCCCATGTCCTTCTGAAAGCACAACCCACCGCAGCAGTCAAT 7180
Db 1441 CCCAAACCACTGTTTCATCCCATGTCCTTCTGAAAGCACAACCCACCGCAGCAGTCAAT 1500

QY 7181 CCAAGCCCCAAGTTGCAGCACAGTCTCAGCCTCAAAGTAATGTCCAAAGGACAGTCTCTTG 7240
Db 1501 CCAAGCCCCAAGTTGCAGCACAGTCTCAGCCTCAAAGTAATGTCCAAAGGACAGTCTCTTG 1560
QY 7241 TTCGTGTCCAAAGTCCATCAGAGTCCGAATAGTCCATCAACTCCCAACTGCTCTC 7300
Db 1561 TTCGTGTCCAAAGTCCATCAGAGTCCGAATAGTCCATCAACTCCCAACTGCTCTC 1620
QY 7301 CTGGCAACAATCCAGAGTTTACAGCTTACAACTCACAACCGATTCCAAATTCACACACATA 7360
Db 1621 CTGGCAACAATCCAGAGTTTACAGCTTACAACTCACAACCGATTCCAAATTCACACACATA 1680
QY 7361 CATCTCTTCAGATACCTTCCCAAGGCCAGCCACAGTCACAAACCCCAAGT 7409
Db 1681 CATCTCTTCAGATACCTTCCCAAGGCCAGCCACAGTCACAAACCCCAAGT 1729

RESULT 10
AK022570
LOCUS
DEFINITION
Homo sapiens cDNA FLJ12508 fis, clone NT2RM2001716, highly similar to Homo sapiens BPTF mRNA for bromodomain PHD finger transcription factor.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

AK022570 2057 bp mRNA linear PRI 01-AUG-2002
Homo sapiens cDNA FLJ12508 fis, clone NT2RM2001716, highly similar to Homo sapiens BPTF mRNA for bromodomain PHD finger transcription factor.
AK022570.1 GI:10434035
oligo capping; fis (full insert sequence).
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,I. and Sasaki,N.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2057)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics-Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
Location/Qualifiers
1..2057
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM2001716"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RM2"
/note="cloning vector: pME18SFL3-mRNA from uninduced NT2 neuronal precursor cells."
BASE COUNT 666 a 566 c 435 g 390 t
ORIGIN

Query Match 17.4%; Score 1719.4; DB 9; Length 2057;
Best Local Similarity 99.7%; Pred. No. 1.8e-265;
Matches 1723; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 5681 TAATTAAGAGGAGAGATGTTGGTCTTATGCGATTCGATTGATTCATCAGAGAAA 5740
Db 1 TAATTAAGAGGAGAGATGTTGGTCTTATGCGATTCGATTGATTCATCAGAGAAA 60

QY 5741 TCATTTCCTCCATTGGAGTTCCAGAACACCAACCAAGAAAGCCCTACACCTCAGAGGAAG 5800
DB 61 TCATTTTCTCCATTGGAGTTCCAGAACACCAACCAAGAAAGCCCTACACCTCAGAGGAAG 120
QY 5801 GCCTTCGATCAAGTGCACCTCGGCCCAAGAGAGACCAAGAAAGCCCAAGCAAACTGGCCCTG 5860
DB 121 GCCTTCGATCAAGTGCACCTCGGCCCAAGAGAGACCAAGAAAGCCCAAGCAAACTGGCCCTG 180
QY 5861 TTATTATTGAACCTTGGGTAGCAGAGAAAGAACTGGAAATTTGGGGAGATCAGGGCAATTTG 5920
DB 181 TTATTATTGAACCTTGGGTAGCAGAGAAAGAACTGGAAATTTGGGGAGATCAGGGCAATTTG 240
QY 5921 CTGAGAGATGGAGAAAGAAAGCAAGCAAGCTTGAGCAACAGGCTAAGAAACAGCTGG 5980
DB 241 CTGAGAGATGGAGAAAGAAAGCAAGCAAGCTTGAGCAACAGGCTAAGAAACAGCTGG 300
QY 5981 AGCAGCAAGAGCCGACAGTATTGCAACTTCCACTTCTCCCAACCAAGCAGTACACCA 6040
DB 301 AGCAGCAAGAGCCGACAGTATTGCAACTTCCACTTCTCCCAACCAAGCAGTACACCA 360
QY 6041 GCACCATCTCTCCAGCAGACAGAAAGTTATGGTGGCCCCCATTAAGTGGCTCAGTTTACAAC 6100
DB 361 GCACCATCTCTCCAGCAGACAGAAAGTTATGGTGGCCCCCATTAAGTGGCTCAGTTTACAAC 420
QY 6101 GAACCAAAATGGTACTTAAGTTGATCTCCAGCTACAGTAACTTCAACAAA 6160
DB 421 GAACCAAAATGGTACTTAAGTTGATCTCCAGCTACAGTAACTTCAACAAA 480
QY 6161 ACAAGAACTTTTCATCAAACTTTGCTACATGGGTTAAGCAAGCCAGTCAAAATTCAGGCG 6220
DB 481 ACAAGAACTTTTCATCAAACTTTGCTACATGGGTTAAGCAAGCCAGTCAAAATTCAGGCG 540
QY 6221 TTGTTCAAGTACAGCAGAAAGTCTCGGGTATCATTCATCAAGTACAGGTACAGTCAAG 6280
DB 541 TTGTTCAAGTACAGCAGAAAGTCTCGGGTATCATTCATCAAGTACAGGTACAGTCAAG 600
QY 6281 AAACCTTTACTTATTCAGCCAGGACACAGCAGTCAAAATTAGGCCCAATACCTCAG 6340
DB 601 AAACCTTTACTTATTCAGCCAGGACACAGCAGTCAAAATTAGGCCCAATACCTCAG 660
QY 6341 GCCTCTGGAGGAACACAGCAATTCACAAGTAATTCACAGGCGCTCAGATTGCGCCTGGTA 6400
DB 661 GCCTCTGGAGGAACACAGCAATTCACAAGTAATTCACAGGCGCTCAGATTGCGCCTGGTA 720
QY 6401 TGACCGTGATTAGAACACCACTCCACAGTCAACACTAGGAAGGCAATTTATTCGAACAC 6460
DB 721 TGACCGTGATTAGAACACCACTCCACAGTCAACACTAGGAAGGCAATTTATTCGAACAC 780
QY 6461 CTGTGATGGTACAGCCAGGTGCTCTCAGCAAGTGATGACTCAAAATCATCAGGGGCGAG 6520
DB 781 CTGTGATGGTACAGCCAGGTGCTCTCAGCAAGTGATGACTCAAAATCATCAGGGGCGAG 840
QY 6521 CTGTCTCCACTGCACTCGCCGCCCTAACACGGTTTCTCTCAACACCTTGGCAGAAAGCT 6580
DB 841 CTGTCTCCACTGCACTCGCCGCCCTAACACGGTTTCTCTCAACACCTTGGCAGAAAGCT 900
QY 6581 TAACCTCAGCAAGCTCCTCAATTAATATACAGTCTTTCAGGCTCACAACCCCTCGCCCCC 6640
DB 901 TAACCTCAGCAAGCTCCTCAATTAATATACAGTCTTTCAGGCTCACAACCCCTCGCCCCC 960
QY 6641 AACCAAGCAAGTCAAGCTCAGCATGCTCAACTTACTCAGTTAACACAGGCGCCAGGTG 6700
DB 961 AACCAAGCAAGTCAAGCTCAGCATGCTCAACTTACTCAGTTAACACAGGCGCCAGGTG 1020
QY 6701 GCATCAAGGTTTGACAGTAGTAAATCAAGCAAGGCTCAAACTTACTTGGACAGTTGCGAGT 6760
DB 1021 GCATCAAGGTTTGACAGTAGTAAATCAAGCAAGGCTCAAACTTACTTGGACAGTTGCGAGT 1080
QY 6761 TGATACCTCAAGGGGTGACTGTACTTCCAGGCCAGCCAGCAGCTAATGCAAGCTCAAA 6820
DB 1081 TGATACCTCAAGGGGTGACTGTACTTCCAGGCCAGCCAGCAGCTAATGCAAGCTCAAA 1140
QY 6821 TGCCAAATGTTACTGTTACGGATTCTCTTTTACCCCATTTGGCAACACAGCCACACAG 6880

DB 1141 TGCCAAATGTTACTGTTTACGCGATTCTCTTTTACCCCATTTGGCAACACAGCCACCAG 1200
QY 6881 CCAGCACCCACACACCACTGTTTCCACGACAGCAGCAGGTACAGGTGAACAAGGCGAGA 6940
DB 1201 CCAGCACCCACACACCACTGTTTCCACGACAGCAGCAGGTACAGGTGAACAAGGCGAGA 1260
QY 6941 GTAACTGTCCACCCAGATCAGGTACATCAAGCAAAACCCCTGCCACAGCTCAGTCAT 7000
DB 1261 GTAACTGTCCACCCAGATCAGGTACATCAAGCAAAACCCCTGCCACAGCTCAGTCAT 1320
QY 7001 CAAGTGTGGTCCAGCAAAACCCAGCAGCAGCTGCTCAGCTTCAAGTTCGCCCCAGC 7060
DB 1321 CAAGTGTGGTCCAGCAAAACCCAGCAGCAGCTGCTCAGCTTCAAGTTCGCCCCAGC 1380
QY 7061 CCCAAACCCAGCCAGTCCGCCAGCTCAGCTGAAGTTTCAAGCTCAGCTTCAAGTTTTCAGA 7120
DB 1381 CCCAAACCCAGCCAGTCCGCCAGCTCAGCTGAAGTTTCAAGCTCAGCTTCAAGTTTTCAGA 1440
QY 7121 CCCAAACCAACTGTTTTCATCCATGTCCTTCTTGAAGCAACACCCAGCCAGCAGTCAT 7180
DB 1441 CCCAAACCAACTGTTTTCATCCATGTCCTTCTTGAAGCAACACCCAGCCAGCAGTCAT 1500
QY 7181 CGAAGCCCAAGTTGCAGCAGCTCTCAGCTCAAAAGTAATGTCCAAAGGACAGTCTCCTG 7240
DB 1501 CGAAGCCCAAGTTGCAGCAGCTCTCAGCTCAAAAGTAATGTCCAAAGGACAGTCTCCTG 1560
QY 7241 TTGCTGTCCAAAGTCCATCAGACTCGAATAGTTCATCAACTCCATCCCAACTGTCTC 7300
DB 1561 TTGCTGTCCAAAGTCCATCAGACTCGAATAGTTCATCAACTCCATCCCAACTGTCTC 1620
QY 7301 CTGACCAACAATCCAGGTTTCAGACTACACCTCACAACCGATTCGAATTCACCAACATA 7360
DB 1621 CTGACCAACAATCCAGGTTTCAGACTACACCTCACAACCGATTCGAATTCACCAACATA 1680
QY 7361 CATCTCTTCAGATACCTTCCCAAGCCAGCAGCTCACAACCCAGGT 7409
DB 1681 CATCTCTTCAGATACCTTCCCAAGCCAGCAGCTCACAACCCAGGT 1729

RESULT 11
BC032289 2549 bp mRNA linear ROD 22-APR-2003
LOCUS Mus musculus fetal Alzheimer antigen, mRNA (cDNA clone
DEFINITION IMAGE:5368380), partial cds.
ACCESSION BC032289
VERSION BC032289.1 GI:21619020
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2549)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Sheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Maman, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shcherchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932

2 (bases 1 to 2549)

Strausberg, R.

Direct Submission

Submitted (03-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGSC) Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs-r@mail.nih.gov

cDNA Library Preparation: The Cepko Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Galithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nhgri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,

Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R.,

Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,

Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 65 Row: h Column: 12

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

Location/Qualifiers

1. .2549

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:5368380"

/tissue_type="Eye, retina, mouse strain C57BL/6"

/lab_host="NIH-MGC_94"

/note="Vector: pCMV-SPORT6"

<1. .2549

/gene="Palz"

/db_xref="LocusID:207165"

/db_xref="MGI:2444008"

<1. .2412

/gene="Palz"

/codon_start=1

/product="Palz protein"

/protein_id="AAH32289.1"

/db_xref="GI:21619021"

/db_xref="LocusID:207165"

/db_xref="MGI:2444008"

/translation="PRVTRPACPDNRSLQVEDMETESPVKRVPSPVTRTGGSN

LSKGFMDNGLPSSKDENVGESORKTIVETWTMTSTVATPESKTVIKVAKDQTVV

STENCAKSTVTTTITTTTKLSTFSPDGVDTISVKEQSKVITVTTVDSLTAGSTL

VTSMTVSKYSTRDRKLMKFSRPKTRSGTALPSYKFTVKSSKSLFVLPNDLKL

LARKGIREVYFNYNAKPAIDWIPSPRPFTFGITWRVLTQVTKSLAGVSLMLRLW

ASLRWDDMAAKAPGGGSTRSTETETITTEIIKRDRVPGVIRSEYCIKIKICPIG

VPEAKPTPQRKGLRSLRALKRPKPTPKOTGPVILESWAEELELEWEIRAFERY

EKEAQAEOQTQRLEQOKPAVTAASTSPANTSSVSPAQKWWAPLSCSVTPGT

KMLVATKSPATVTFQONKPHOTFAWVKQGSNSGVQVQQRVLGIIIPSTGTSQ

QTFSTFQRTATVTRPNTSAGTTTSQVITGPQIRPMTVTRIPQALGKAI

RPVVPQGVQKVVQTIIRQPGVSTALSPASSAPVKQGLTPGAAGPLQPSAPH

SPRQOQVQLTMAQLTQTOGHGNOGLTVVIOGOGTTGLOLIPGMYLPGPGO

OLMQAAMPNGTVQRFLETPLSATASSSSSSNSTTNTAAGSGEOKSKILPQTQ

VOPATTLAPTSSSVSPAEOQPQAPAAQPQPPQPPQPPQPPQPPQPPQPPQPPQ

PSLLS"

FEATURES
source

gene

CDS

BASE COUNT 808 a 711 c 583 g 447 t
ORIGIN

Query Match 17.4%; Score 1714; DB 10; Length 2549;

Best Local Similarity 82.8%; Pred. No. 1.3e-264;

Matches 1998; Conservative 0; Mismatches 380; Indels 36; Gaps 2;

QY 4767 CACACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAGATATGGAAT 4826

Db 12 CACACCCCGGCCCTTGTCCAGATCGGAATCACTTAGTCAGGTAGAGATATGGAAC 71

QY 4827 AGAAACCTCAGAAAGTTAAGAAAGTTACTTCATCACCCTATTACTTCTGAAGAGAAATCTAA 4886

Db 72 TGAAGAGTCAGAGGTTAAGAGAGTTATCCGCTCTCTGTTAGAACTGGAGAGGGCTCTAA 131

QY 4887 TCTCAGTAATGACTTTATTGATGAAATGGTCTGCCCATCAACAAAAATGAAAAATGCTCAA 4946

Db 132 CTTAGCAAGGGCTTCATGGATGACATGGACTGCCTCCAGCAAGAGCAGACGCTCAA 191

QY 4947 TGGAGAAATCTAAAAAGAAACCGTCATCACAGAAAGTCACAGATGACCTCCACAGTGGC 5006

Db 192 TGGAGAAATCTAAAAAGAGAGCGGTATCATCAGAAAGTCACCAATGACATCTACTGTTGC 251

QY 5007 CACAGAAATCAAAACTGTGATCAAGTAGAAAGGCAAGCAAACTGTGTTCTTC 5066

Db 252 CACAGAAATCAAAACTGTGATCAAGTAGAAAGGCAAGCAAACTGTGTTCTTC 311

QY 5067 CACAGAAATCTGCAAAATCCACTGTCTCAACACCACTACCAAGTGCACCAAGCTTC 5126

Db 312 CACAGAAATCTGCGCAGATCCACAGTCCAACTACCCTACCATCTGTGACAAAGCTCTC 371

QY 5127 CACACCCCTCCACAGCGCGAGTGTGACATCATCTCTGTAAAGAGCAGAGCAAAACCT 5186

Db 372 CACACCCCTCCACAGACACCGGTGTGATACCATCTCTGTAAAGAGCAGAGTAAACCT 431

QY 5187 GGTCAACCGCAGTGCACAGACTCCTGTACCAACCGGAGGAGGACACTGTTTACATCTAT 5246

Db 432 GGTCAACCGCAGTGCAGGACTCCTGTGACCACTGCAGGAAGCAGACTGTTGACATCTAT 491

QY 5247 GACTGTGAGCAAGAGTATTCCACAGAGCAAGAACTGAACTGATGAAATTTTCAAGACC 5306

Db 492 GACCTGTGACCAAGAGTACTCCACAGAGACAGAGGTTAAAGCTGTGAAAGTCTCAGACC 551

QY 5307 AAAGAAGACTCGTTCAGGTACAGCTCTGCCATCTATAGAAATTTGTTTACCAAGAGCAC 5366

Db 552 CAAGAAGACCCCTCAGGCACCTGCCCTCTATAGAAAGTTCGTTTACCAAGAGCAG 611

QY 5367 CAAGAAGACATTTTGTGTTTGTGCTTAATGATGACTTAAAGAGTTGGCCCGGAAAGGAGG 5426

Db 612 CAAGAAGACATTTTGTGTTTGTGCTTAATGATGACTTAAAGAGTTGGCCCGGAAAGGAGG 671

QY 5427 AATCCGAGAGGTCCTTATTTTAAATACAAATGCAAAACCTGCTTTGGATATATGGCCATA 5486

Db 672 GATCCGGAGGTCCTTACTTTAATTAACCGGAAACCTGCTTTGGATATATGGCCATA 731

QY 5487 TCCTTCTCCTAGACCGACCTTTGGCATCACTTTGGAGGTATAGACTTCAGACAGTAAAGTC 5546

Db 732 CCCTTCTCCTAGACCGACCTTTGGTATCATCATGGAGGTATAGACTACAGACGTTAAAGTC 791

QY 5547 CTTACTGTGAGTACCCCTGATGTTACGTTTACTGTGGCAAGTTTTCAGATGGATGATAT 5606

Db 792 CTTGCTGGAGTACCCCTGATGTTACGTTTGTGTTGCGCCAGTTCGATGGATGGATGACAT 851

QY 5607 GSCGCCCAAGGTTCTCCAGGAGGAGGTAGTAGCAGCAGAGAAACATCGAAACCTGAAAT 5666

Db 852 GSCGCCCAAGGTTCTCCAGGAGGAGGTGCTACCGAGCAGAAACATCTGAAACCTGAAAT 911

QY 5667 CACAAACACAGAAATAATTAAGAGGAGAGATGTTGTCCTTATGGCATTTCGATTGGAATA 5726

Db 912 CACAAACACTGAGATAATTAAGAGCGGTGATCTGGGCCCTTATGGCATTTCGTTCTGAGTA 971

QY 5727 TTGTATCAGGAAATCATTTTGTCCCATTTGGAGTTCCAGAAACCAACCAAGAACCGCTTAC 5786

Db	972	TTGTATCAGGAAGATCATTTTGCCTCCCATTTGGAGTTTCAGAAAGCACCAAAAGAAACACCCAC	103
Qy	5787	ACCTCAGAGAAAGGCCTTCGATCAAGTCACACTGCGGCCCAAGAGACACAGAAACGCCCAA	5846
Db	1032	ACCTCAGCGGAAGGGCTCCCGATCAAGTCGACATTGCGCCAAAGAGACACAGAGACCCCAA	1091
Qy	5847	GCAAACCTGGCCCTGTATTATTTGAAACCTTGGGTAGCAGAGAGAAGAACTGGAATTTGGGA	5906
Db	1092	GCAGACGGGCCCTGTTATCATTTGAGAGCTGGGTAGCAGAAAGAGAGCTGGAGCTCTGGGA	1151
Qy	5907	GATCAGGGCATTTGCTTGAGAGAGTGGAGAAAGAAAAGGCCAAAGCAGTTGAGCAACAGGC	5966
Db	1152	GATCAGGGCTTTGCTTGAGAGAGTGAGAAAGAAAGGCTCAAGCAGCAGCAGCAGAC	1211
Qy	5967	TAGAAACAGACTGGAGCAGCAGAGCGGACAGTGATTGCAACTTCCACTACTTCCGCCAAC	6026
Db	1212	CAAGAAACGTTTGGAGCAGCAGAGCGCTCGAGTCATCGCAGCGTCCACCACTGCCCGAGC	1271
Qy	6027	AAGCAGTACAACAGCAGCACCACTCTCTCCAGCACAGAAAGTTATGGTGGCCCCCATAAAGTGG	6086
Db	1272	AAACAACAAGCAGCACTGTCTCTCAGCCCGAAGGTCGATGGTGGCTCCCTTTAAGTGG	1331
Qy	6087	CTCAGTTTACAACCTGGAAACAAAATGGTACTAACTACTTAAAGTTGGATCTCCAGCTACAGT	6146
Db	1332	CTCGCTCACCCCTGGAACAAGATGGTGTCTGGCCACAAAAAGTCGGCTCGCGCTGCCACAGT	1391
Qy	6147	AACATTTCAACAACAACAGAACTTTCATCAAACTTTGCTACATGGGTTAAGCAAGGCCA	6206
Db	1392	CACATTTCCAGCAAAACAAGAACTTCCATCAAACTTTGCCACGTGGGTGAAGCAAGGCCA	1451
Qy	6207	GTCAAAATTCAGGCGCTTCTCAAGTACAGCAGAAAGTCTCGGTGTATCATTCATCAAGTAC	6266
Db	1452	GTCCAATTCAGGCATGTCCAGGTACAGCAGAAAGTCTCTGGGATCATTTCCATCGCACTAC	1511
Qy	6267	AGGTACACAGTCAAGCAAAACCTTTACTTCAATTCACAGCCACGACACACAGTCAACAATTAG	6326
Db	1512	AGGTCCAAGCCAGCAAAACCTTTACTTCATTCACGCCACGAGACTTCAAACTGTTCACAATTAG	1571
Qy	6327	GCCCAATACCTCAGGCTCTCGAGGAACCAAGAAGAAATTCACAAGTAATACAGGGGCTCA	6386
Db	1572	GCCCAATACCTCAGCCTCTCAGGAACCACTACCAATCAGAGTGATCAGAGGCTCA	1631
Qy	6387	GATTCGCCCTGGTATGACCGTGATTAGAACACCACTCCACAGATCAACACTAGGAAGGC	6446
Db	1632	GATCCGCCCTGGCATGACAGTGATTAGAAGCCCACTCCAGCAGCAGCATCGGAAAGGC	1691
Qy	6447	AATTTATCGAACACTCTGTATGTFACAGCAGAGTGTCTCTCAGCAAGTGATGACTCAAAAT	6506
Db	1692	GATCATTTGCAACACCATGGTGGTGACAGCCAGGCACACCCCAACAAGTTGTGACTCAAAAT	1751
Qy	6507	CATCAGGGGGCAGCTGTCTCCACTGCACTGCTCGGCCCTTAACACAGGTTTTCCTCAACACC	6566
Db	1752	CATCCGAGGACAGCCTGTTTCCACTGGCATCTCTGTCTCAAGTACAGCTTCTCTCAGACCC	1811
Qy	6567	TGGCAGAAAAGCTTAATTCAGCAACAGTCCACTTCAAAATATACAGTCTTCAGCCTCACA	6626
Db	1812	TGTACAAAAGGGCTTAACTCCAGAGCAGGGGCTGGGGCTCTGAGGCTTTCAGCCCCACA	1871
Qy	6627	ACCCCTCTCGCCCCAACAAAGGACAAGTGAAGCTCACCATGGCTCAACTTACTCAGTTAAAC	6686
Db	1872	CAGCCCTCGTCTCAGCAAGGGCAGGTGAAGCTCACCATGGCTCAGCTCACTCAGCTCAC	1931
Qy	6687	ACAGGGCCACGGTGGCAATCAAGGTTTTGACAGTAGTAAATTCAGGACAAAGTTCAACTAC	6746
Db	1932	GCAGGGCCATGGCGGCAACCAAGGTTTTAACAGTAGTAGTTACAGGACAAAGGGCAGACTAC	1991
Qy	6747	TGGACAGTTTGAGTTGATACCTCAAGGGGTGACTGTACTCCAGGCCACCGCCAGCAGCT	6806
Db	1992	TGGACAGTTTGAGTTGATACCTCAGGGGATGACAGTCTCTCCAGGCCAGGCAGCAGCT	2051
Qy	6807	AATGCAAGCTGCAATGCCAAATGTTACTGTTTACGGGATTTCTCTTTTACCCCATTTGGCAAC	6866
Db	2052	CATGCAAGCTGCAATGGCGAATGGCAGCGTTTACGCGTTTCTCTTTTACCCCATTCACAC	2111

Qy	6867	AACAGCCACACACCCAG-----CACCAACCAACCACTGTTTCCACGACGC	6914
Db	2112	ATCAGCCACGCCGCCACGACGACGACGACGACCAACAGCAGCAGCACTACTAATGCTACAGC	2171
Qy	6915	AGCAGGTACAGGTGAACAAAGGACGAGTAAACTGTCTACCCCGACATGCGAGGTACATCAAGA	6974
Db	2172	AGCAGGCTCAGGGAACAAACACAGAGTAAATATTACCCGACACCCAGGTACACCCAGC	2231
Qy	6975	CAAAACCCCTGCCACACGCTCAGTCAATCAAGTGTGGTCCAGCAAAAGCCGACCCACAGAC	7034
Db	2232	CACAACCCCTGGCACCAACTCAGTTCATCAAGTGTGAGTCTGTCGACGAAGCCGCCACGACCC	2291
Qy	7035	TGCTCAGCCTTCAGCTCGGCCCGCCAGCCCAACCCAGCCGCCAGTCCCGACGCTCAGCCCTGA	7094
Db	2292	TGCTCAGCCTTCAGCACAACCCCGACCCCGCCAGCCCGCC-----	2331
Qy	7095	AGTTTCAGACTCAGCCTGAAGTTCAGACCCCAAACTGTTTCATCCCATGTCCTTCTGA	7154
Db	2332	---CCAGCTCAGCCTGAAGTTCAGACCCCGACCCAGCAGTGTCTGTCCTCATGTCCTTCTGA	2387
Qy	7155	AGCACACCCACCC 7168	
Db	2388	AACACAGCCCTCCC 2401	
RESULT 12			
BC046973		2769 bp	linear
LOCUS		2769 bp	linear
DEFINITION		Mus musculus hypothetical protein LOC268496, mRNA (cdna clone	
ACCESSION		MG:54978 IMAGE:6487444), complete cds.	
VERSION		BC046973	
KEYWORDS		BC046973.1 GI:28422395	
SOURCE		MGC.	
ORGANISM		Mus musculus (house mouse)	
REFERENCE		Mus musculus	
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
		1 (bases 1 to 2769)	
		Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausch, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffen, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Vittal, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.	
		human and mouse cDNA sequences	
		generation and initial analysis of more than 15,000 full-length	
		human and mouse cDNA sequences	
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
MEDLINE		22388257	
PUBMED		12477932	
REFERENCE		2 (bases 1 to 2769)	
AUTHORS		Strausberg, R.	
TITLE		Direct Submission	
JOURNAL		Submitted (14-FEB-2003), National Institutes of Health, Mammalian	
		Gene Collection (MGC), Cancer Genomics Office, National Cancer	
		Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,	
		USA	
REMARK		NIH-MGC Project URL: http://mgc.nci.nih.gov	
COMMENT		Contact: MGC help desk	
		Email: cgabbs-r@mail.nih.gov	
		Tissue Procurement: Gilbert Smith, Ph.D.	

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hri.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lario, P., Legaspi, R.,
 Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>
 Series: IRAK Plate: 100 Row: m Column: 22
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Genomescan gene
 prediction, Similarity but not identity to protein.

FEATURES

source

1. .2769
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N-3"
 /db_xref="taxon:10090"
 /clone="MGC:54978 IMAGE:6487444"
 /tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
 old mouse. Taken by biopsy."
 /clone_lib="NCI_CGAP_Mam2"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1. .2769
 /gene="LOC268496"
 /db_xref="LocusID:268496"
 163. .2577
 /codon_start=1
 /product="LOC268496 protein"
 /protein_id="AAH4973.1"
 /db_xref="GI:28422396"
 /db_xref="LocusID:268496"
 /translation="MVSEDEEEEDGDAEETQDSEDEEDDEEDDDSDYPEEMED
 DDDASYCTESSRSHSYSTPGRRKPRVHRSPILKEEKDIPLEFPKSEDLMP
 NEHIMNIAIYEVNRFGNVLRLSPFCFEDCAALYSQECQLMAEMHVALKLYLRE
 EDTNITTPGADLKDSVNSTLYFIDGMTPELVRCSDXKEYHVLVPOEAEDYPYG
 PVENIKVLQFLVDQFLTNIAREELMSGVQYDDHCRVCHKLDLCCETCSAVIH
 LECVPPLLEVPEDSEQCEVCVAHKVPQVTDCAVEOKKPYVHEPTGYDRSRKYH
 FLNRLIIEEDTDNEKKVWYISTVKQLAELIDLDKGWEALCRVLDEIREMQY
 HMDVTEDLTNKARGSNKSFLLAAANBEIILDSLRKRGEDIDCDSPEDPEKDKHEGN
 SSKDAEKREEADPSADKDADSKGLEEPGHGKPEEDSNRTCDSSNTSATTSQPN
 LFTCSSELTSSQSDSAKAADPEIGERDSHTPVSVHREEIGDFLEKSNSEYSEPGA
 GKTSGSRIIPRLNRPESKLSQAKQOAAAAHBANKLFKEGKEVLVNVSGSEVSL
 SPYKKEVMKGNINNYFKLGOEGRYVHNOYSTNFKLNKHQHRDHDKRHLAHKFC
 LTPAGEFWNGSVHSGKVLITSLRLTIQLESNIPSSFLHFNWASHRANWIKAVQMC
 SPRFELALALECAVPVMLPIWRESLGHRLHRTMSIEREKEKVKYKKKKKGR
 PFFFF"

BASE COUNT 862 a 609 c 753 g 545 t
 ORIGIN

Query Match 17.3%; Score 1708.4; DB 10; Length 2769;
 Best Local Similarity 80.0%; Pred. No. 1e-263;
 Matches 2193; Conservative 0; Mismatches 336; Indels 213; Gaps 7;

QY 96 CAGCACCAGCCCGGGGGGGAGAGCGGGGGCGGCACGACGGGGGGG 155
 |||||
 Db 6 CAGCGCCAGCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 65
 QY 156 GGGCGGGCGGCCACCTGTCTCCCGACACCGCCGGCCGGAGGGCGCTCAACAAAGTGTG 215
 |||||
 Db 66 G-----GCGGCCACCTGGCCCGGACACACCCCGCCCGCGCGTCAACAAGTGTG 119

QY 216 GTACGATACACGAGAGCGAGCGCGTGGAGGAAGAGGAGGACATCGTCTCCAG----- 270
 Db |||||
 QY 120 GTACCGTACCCACGAGAGTGCACGACGACGAGGAGGAGGACATGCTCGGAGGAGGA 179
 Db |||||
 QY 271 -GAGGAGGAGGAGGAGGAGCGGACCGCGAGGAGACCCAGGATTCCTGAGGACACGAGGA 329
 Db |||||
 QY 180 CGAGGAGGAGGAGGAGGATGGCGACCGCGAGGAGACCCAGGATTCGAGGACGAGGAGGA 239
 Db |||||
 QY 330 GGATGAGATGGAGAGGAGGACGACGATGACTCCCGATTCGCGGAGGAGATGGAAGACGACGA 389
 Db |||||
 QY 240 AGATGACATGGAAGAGACGACGATGACTCCCGATTCGCGGAGGAGATGGAAGACGACGA 299
 Db |||||
 QY 390 CGACGACGCCAGTTACTGACGAGAAAGACAGCTTCAGAGGCCATAGTACCTACAGCAGCAC 449
 Db |||||
 QY 300 CGACGACGCTGTTACTGACGAGAAAGACAGCTTCAGAGGCCATAGTACATACAGCAGCAC 359
 Db |||||
 QY 450 TCCAGGTAGCGGAAACCAAGAGTACATCGGCTCGTCTCCCTATATTTGGAACAAAAAGA 509
 Db |||||
 QY 360 TCAGGTAGCGGAAACCAAGAGTACATCGGCTCGTCTCCCTATATTTGGAACAAAAAGA 419
 Db |||||
 QY 510 CATCCCGCCCTTGAATTTCCCAAGTCTCTGAGGATTTAATGGTGCCTAATGAGCATAT 569
 Db |||||
 QY 420 CATCCCAACCCCTTGAATTTCCGAAGTCTCTGAGGATTTAATGGTGCCTAATGAGCATAT 479
 Db |||||
 QY 570 AATGAATGTCATTCGCCATTTACGAGTACTCGGAACTTTTGGCACTGTTTGGAGATATTC 629
 Db |||||
 QY 480 AATGAATGTCATTTGCCATTTATGAGTGGTGGCGAACTTTTGGCAAGCTTTTAAAGTTGTC 539
 Db |||||
 QY 630 TCCTTTTCCGCTTTGAGGACTTTTGTGACGCTCTGTGAGCAAGACAGCAGTGCACACTCAT 689
 Db |||||
 QY 540 CCTTTTTCCTTTGAGGACTTCTGTGCGGCTCTGTGAGCAAGACAGTGCACACTCAT 599
 Db |||||
 QY 690 GGCAGAGATGCATGTTGTGCTTTTGAAGCAGTCTTCGCGTGAAGAGACAGCACTTCCCAATAC 749
 Db |||||
 QY 600 GSCGGAGATGCACGTGCGCTTCTGAAAGCAGTCTTCGAGAGAGAGACACGCTTACAC 659
 Db |||||
 QY 750 TACCTTTGGACCTGCTGATCTGAAAGATAGCGTTAATTCACACTGTTATTTATAGATGG 809
 Db |||||
 QY 660 CACCTTTGGCCCTGCGGACTTGAAGACAGTGTAAACTCCACCCTGTATTTTCATCAGCG 719
 Db |||||
 QY 810 GATGAGTGGCCAGAGGTGCTGCGGTGCTGAGAGTGTGTAAGAGTGAAGAGTACCATCAGCT 869
 Db |||||
 QY 720 GATGAGTGGCCGAGGTCTGAGGTCTGAGGTCTGAGAGTGAAGAGAGTACCACCATGT 779
 Db |||||
 QY 870 TCTTCCTTACCAGGCGAGAGTACCCATATGGACCAGTAGAGAACAAAGATCAAGT 929
 Db |||||
 QY 780 CCTTCCTTACCAGGCGAGAGTATCCCTACGACCACTGAGGAGACAAATCAAGT 839
 Db |||||
 QY 930 TCTACAGTTTCTAGTCGATCAGTTCTTACACAAATATGCTCGAGAGGAATTTGATGTC 989
 Db |||||
 QY 840 TCTGCAAGTTTCTAGTCGATCAGTTCTTACGACCAATATTTGCTGAGGAGCTGATGTC 899
 Db |||||
 QY 990 TGAAGGGGTGATACAGTATGATGACCATTTGTAGGGTTTGTCAACAACTTTGGGATTTGCT 1049
 Db |||||
 QY 900 TGAAGGGGTAACTCCAGTATGACGACCATTTGTAGGGTCTGTCAACAACTTTGGGATTTGCT 959
 Db |||||
 QY 1050 TTGCTGTGAGACATGTTTACGAGTATACCATTTGGAATTTGTGTAAGCCACCTTCTGAGGA 1109
 Db |||||
 QY 960 GTGCTGTGAGACGTGCTCGGCTGTGTACCATTTAGAGTGTGTGAAGCCGCTCTTGAAGA 1019
 Db |||||
 QY 1110 GGTGCCAGGAGGAGTGGCAGTGTGAAGTCTGTGTAGCACACAAAGTGCCTGGTGTGAC 1169
 Db |||||
 QY 1020 AGTGGCGGAGGAGTGGCAGTGTGAAGTCTGTGTGGCACACAAAGTGCCTGGTGTGAC 1079
 Db |||||
 QY 1170 TGACTGTGTGTGAAATCCAAAAATAAACCATATATTTCCGACATGAACCTATTGCGATA 1229
 Db |||||
 QY 1080 TGACTGTGTGTGAGTCCAAAGANATAACCTACGTTCCGACAGCAACCTATTGGGTA 1139
 Db |||||
 QY 1230 TGATGAAGTCCGAGGAAATACTGGTTCTTGAACCGAAGACTCATATTAAGAAGAGATAC 1289
 Db |||||
 QY 1140 TGACAGAGCCCTAGGAAATACTGGTTTAAACCCGAGACTCATATAAGAAGACAC 1199
 Db |||||
 QY 1290 AGAAATGAAATGAAAGAAAAATTTGGTATTACAGCACAAAGGTCCCACTTGCAGAAAT 1349

Db	1200	AGACAATGAAAAACGAGAAAGTTTGGTACTACAGCAAAAGGTCACGTTAGCAGAGCT	1259
Qy	1350	AATTTGACTGCTAGACAAAGATTTATTTGGGAAGCAGAACTCTGCARAAATTTCTAGAAGAAAT	1409
Db	1260	AATAGACTGGCTTAGACAAAGGATACTTGGGAAGCAGAACTGTGCAGAGTTCTTAGAAGACAT	1319
Qy	1410	GCCTGAAGAAATCCACCCGACACATGACATAACTGAAGACCTGCACCAATAAGGCTCGGG	1469
Db	1320	TCGGGAGGAGATGCAGCAGCATATGGATTAACGGGAAGACCTGACCAACAAGGCTCGTGG	1379
Qy	1470	CAGTAACAATCCTTTCTGGCGCACCTAATGAAGAAATTTTGGAAATCCATAGAGCCAA	1529
Db	1380	CAGCAACAAGCTCCTCTAGCGCACCTAATGAGGAAATTTTGGACTCCTTAAAGATCAA	1439
Qy	1530	---AAAGGGAGACATTGATACTTAAAGCCAGAGAAACAGAAAAAGACAAAGAATGA	1586
Db	1440	ACGAGGGGAAGATTTGACTGTGTCATCAGAGCCCAAGACCCAGAAAGACAAAGCATGA	1499
Qy	1587	GACTGAGAATGA---CTCTAAAGATGCTGAGAAAAACAGAGAAGAAATTTTGAAGACAGTTC	1643
Db	1500	GGGCGAGAACACAGCTCCCAAGATGCTGAGAGAGCAGGGAGGAGCGGAGACCTTTC	1559
Qy	1644	CCTTGAAAAGACAGTCAGCACAAAACACAGATGATGACCTTGACCAAGGAAAATCTGA	1703
Db	1560	CGCAGACAGGAGCGCTGACACAAAGGCCCTGGAGGAAGAGCCCGGCGCATTGGAAAACCTGA	1619
Qy	1704	GGAGCCCAACAGAAGTTGGGGATAAAGGTAACTCTGTGTCAGCAAAATCTTTGGGAGACAACAC	1763
Db	1620	G-----	1620
Qy	1764	AACAAATGCAACTTCAGAAGAGACTAGTCCCTCTGAAGGAGGAGGCCCTGTGGGTGTCT	1823
Db	1621	-----	1620
Qy	1824	CTCAGAAAACCCCGATAGCAGCAACATGGCAGAGAAGGTTGGCATCTGAGCTCCGCCA	1883
Db	1621	-----	1620
Qy	1884	GGATGTCGCAAGAAGAACCTTAACAAGACATGTGAGAGCAGTAACACTAGTGTCTACCACTAC	1943
Db	1621	-----GAAGACTCTAACAGGACATGTGACGACAGTAACACTAGTGTCTACCACTGC	1670
Qy	1944	CTCCATCCAGCCTAATCTGAAAAACAGTAACACAGCAGTGAACATAAATTTCTTCCAGAG	2003
Db	1671	CTCAGCCAGCCTAACCTGGAGACCTG---CAGCAGCAGCAGCTCACITCTTCTCCAGAG	1727
Qy	2004	TGAATCTGTAAGCGACTGATGATCCTGAAAATGGAGAAAGAGATCTCATACACTGT	2063
Db	1728	TGACTCTGCTAAGCAGCGGATGACCCCTGAGATTGGAGAAAGAGACTCCCATACACTGT	1787
Qy	2064	CTCTATTCCAGGAAGATAGTAGTGTATTTACATCGGAGAAGTCCACCCGGGAGGTAAAG	2123
Db	1788	CTCTGTTACGAAGAGA---TAGTGTATTTTCAGNCTGGAGAGTCCAAATGGAGAAGTTTC	1844
Qy	2124	TGAATCTCTGGAGCTGGAAAAGAGGACATCTGGCTCAACTTCGAATCATACCAAGATTGCG	2183
Db	1845	CGAGTCCCTCGGAGCTGGGAAAAGAACATCTGGGTCCACTCTGATCATCACAGACTGAG	1904
Qy	2184	GAATCCAGATAGCAAACTTAGTACAGCTGAAGAGCCAGCAGGTGGCAGCCGCTGCACATGA	2243
Db	1905	GAACCCGGAGCAAACTCAGTACGCTCAAGAGTCACAGGTGGCAGCTGCAGCCCATGA	1964
Qy	2244	AGCAAAATAATTTATTAAGGAGGCCAAGAGGTACTGGTAGTTAACTCTCAAGGAGAAAT	2303
Db	1965	AGCAAAATAATTTTAAGGAGGGAAAAGAGGTCTGGTGGTTAACTCCCAAGGAGAGGT	2024
Qy	2304	TTACGGTTGAGCACCAAAAAGGAGGTGATCATCAAGGAAATATCACAATATTTTAA	2363
Db	2025	TTACGGCTTGAGCACCAAAAAGGAGGTGGTCATGAAAGGGAACATCAACAACTATTTTAA	2084
Qy	2364	ATTGGGTCAAGAAGGGAAGTATCCGCTCTACCAAACTCAATFACTTCCACCAATTCATTTGC	2423

Db	2085	GCTGGGACAGAAGGAAGTATCGAGTCTACCAATCAATCTACGACCAATTCATTTCG	214
QY	2424	TTTGAATAAGCACACGACACAGAGAAGACCATGATAGAGAAGGCATCTTGCACATAAGTT	2483
Db	2145	TCTGAATAAACACCCAGCACAGAGAAGACCATGATAGAGAAGGCATCTTGCACATAAAT	2204
QY	2484	CTGCTGACTCCAGCAGGAGAGTTCAATGGAACGGTCTGTGTCATGGGTCCAAAGTTCT	2543
Db	2205	CTGCTGACTCCAGCAGTGGGAGTTCAATGGAACGGTCTGTGTCATGGGTCCAAAGTTCT	2264
QY	2544	TACCATATCTACTGTGAGCTGACTATCATCCCAATTAGAAAAACAACATCCCTTCATCCTT	2603
Db	2265	TACATCTCAACCTGAGACTCACCATCACACAGCTAGAGCAACATCTCTTCATCCTT	2324
QY	2604	TTTTTCATCCCACTGGGCATCATATAGGCAAAATTTGGATCAAGCACTTCAGATGTGTAG	2663
Db	2325	CCTTCCACCCCAATTTGGCTTTCACACAGGCGTAAATTTGGATCAAGGCGCTTCAGATGTGTAG	2384
QY	2664	CARACCCAGAGAATTTTCATTTGGCTTTAGCCATTTTCGGAGTGTGCAGTTTAAACCAAGTTGT	2723
Db	2385	TAAACCCAGAGAGTTTCGGTTGGCTCTCGCAATTCCTGGAGTGTGCAGTTTAAACCAAGTTGT	2444
QY	2724	GATGCTACCAATATGGCGAGAATTTTGGAGCATATACCGAGTTACACGGGATGACATCAAT	2783
Db	2445	GATGCTCCCAATATGGCGGAGTCTCTAGGACATACCGGTTACATAGGATGACATCAAT	2504
QY	2784	TGAAGAGAAGAAAGAGAAAGTCAAAAAAGAGAGAAGA 2825	
Db	2505	TGAAGGGAAGAAAAGAGAAAGTTAAAAAAGAGAGAGAAA 2546	
RESULT 13			
BC037661			
LOCUS		2481 bp	linear
DEFINITION		Mus musculus hypothetical protein LOC268496, mRNA	ROD 16-APR-2003
ACCESSION		IMAGE:5353895, partial cds.	
VERSION		BC037661	
KEYWORDS		BC037661.1 GI:22902292.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	
AUTHORS		1 (bases 1 to 2481)	
		Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Narusina,K., Farmer,A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J., Abramson,K.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,R.J., Mallek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickinson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,X.S., Krzywinski,M.I., Skalska,U., Smallos,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.	
TITLE		Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
MEDLINE		22388257	
PUBMED		12477932	
REFERENCE		2 (bases 1 to 2481)	
AUTHORS		Strausberg,R.	
TITLE		Direct Submission	
JOURNAL		Submitted (12 SEP 2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive Room 11A03, Bethesda, MD 0892-2590,	

REMARK
COMMENT

USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue procurement: Jeffrey Green M.D.
cDNA Library preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Madduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 81 Row: n Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: Genomescan gene
prediction.

FEATURES
source

Location/Qualifiers
1. .2481
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5353895"
/tissue type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .>2481
/gene="LOC268496"
93. >2481
/db_xref="LocusID:268496"
/gene="LOC268496"
/codon_start=1
/product="LOC268496 protein"
/protein_id="AAH37661.1"
/db_xref="GI:27370657"
/db_xref="LocusID:268496"
/translation="MVSEDEEEEDGAEETQDSEDEEDDEDDDDSDYPREMED
DDDACTESFSRSHSTYSPGRRKPRVHRSPVILEEKDIPPEPKSSEDLMP
NEHMNVIAIYEVVRNFGNVLSPFCFDEFCAALVSOEQTLMAMHVALIKVLR
EDTSNTTGPADLKDVSNTLYFIDGMTPEVLRYCSEKDEYHVLVPOBAEDYPY
PVENKIKVLPVLDQFLVTNIARELNSEGVIOYDDHCRVCKLGLDLCCETCSAVYH
LECVKPPLEPVEDWQCEVCAHKVPVTCVAEVQNKPYVHPHPIGIDRSRKKYH
FLNRLLIIEEDTDNEKKWYISTKVOLAELIDCLDKYWEAELCRVLEDIREMQQ
HMVTEIDLTKARNSKFLAAANEIILSRIRKGEDIDCDSPEDPKDKHEGNN
SSDAEKREAEADP SADKDSKLEEPHGKPEETVEVDKGNVSPANLGNNTN
ASPEETSPCDKSPGCLSETHDSMAEKKVASELPDPVDFRLEKSNSESVESPG
AGKTSGSTRITLRLNPESKLSQLKSOVAAAHEANKLEKEGKVLVNSQEVSR
LSTKKEVYKNINNYFKLQEGKRYRTHNQYSTNSFALNKHQHEHDKRHLAHKF
LCTPAGFPKWSVHSGKVLITSLRLITQLESNIPSSFLHPNWNASHRANWIKAVQM
CSKPRFALALAILCAVKKPVMLPIWRESLGHTRLHRMTSTIEREKKKKKKKKK"

BASE COUNT
ORIGIN

Query Match 17.1%; Score 1690.2; DB 10; Length 2481;
Best Local Similarity 80.0%; Pred. No. 8.5e-261;
Matches 2131; Conservative 0; Mismatches 328; Indels 204; Gaps 4;
176 CCCGACACCGCGCCGGAGGCGGTCAACAAGTGGTACCATCACCACGAGAGCG 235
10 CCGGACACCGCGCCGGCGCGGTCAACAAGTGGTACCATCACCACGAGAGTG 69

QY 236 AGCGCGTGGAGGAAGAGGAGGACATGGTCTCCGAG-----GAGGAGGAGGAGGAGCG 289
Db 70 ACGACGACGACGAGGAGGAGGACATGGTCTCGGAGGAGGACGACGAGGAGGAGGAGGATG 129
QY 290 GCGACGCGGAGGAGGAGGAGGAGGATTCGAGGAGGAGGAGGAGGATGAGATGGAGAGGAGCG 349
Db 130 GCGAGCGCGGAGGAGGAGGAGGATTCGAGGAGGAGGAGGAGGAGGATGAGATGGAGAGAGCG 189
QY 350 ACGATGACTCCGATTATCCGAGGAGGAGTGAAGAGACGACGACGACGACGACGCTTACTGCA 409
Db 190 ACGATGACTCCGATTATCCGAGGAGGAGTGAAGAGACGACGACGACGACGCTTACTGCA 249
QY 410 CGGAAAGCAGCTTCAGGAGGCCATAGTACTACAGCAGCAGCTCCAGGTAGGCGAAACCAA 469
Db 250 CGGAAAGCAGCTTCAGGAGGCCATAGTACTACAGCAGCAGCTCCAGGTAGGCGAAACCAA 309
QY 470 GAGTACATCGGCTCGTCTCTCTATATTGGAAGAAAAGACATCCCGCCCTTGAATTC 529
Db 310 GAGTACATCGGCTCGTCTCTCTCAATATTAGAAGAAAAGACATCCCGCCCTTGAATTC 369
QY 530 CCAAGTCTCTGAGGATTTAATGGTGCCTAATGAGCATATATGAATGATTCATTCGATTT 589
Db 370 CCAAGTCTCTGAGGATTTAATGGTGCCTAATGAGCATATATGAATGATTCATTCGATTT 429
QY 590 ACGAGTACTCGGAACTTTGGCACTGTTTGGAGATTTCTCTCTTTCGCTTTGAGGACT 649
Db 430 ATGAGGTGTGCGGAACTTTGGCAACGTTTTAAGGTTGTGCCCTTTTCTTTGAGGACT 489
QY 650 TTTGTGACGCTCTGAGGAGCCAGGAGCAGTGCACACTATGCGCAGAGATGATGTTGTC 709
Db 490 TCTGTGCGGCTCTGAGGAGCCAGGAGCAGTGCACACTATGCGGAGGATGCGACGTGGCGC 549
QY 710 TTTTCAAAGCAGTCTCGGTGAAGAGACACTTCCAACTACTACCTTTGGACCTGCTGATC 769
Db 550 TTCTGAAAGCAGTCTTCGAGAGAGGAGACACTCTAACACCACTTTGGCCCTCGGACT 609
QY 770 TGAAGATAGCGTTAATTCACACTGTATTTTCATAGATGGGATGACGTGGCCAGAGTGC 829
Db 610 TGAAGACAGTGTAAACTCCACCTGTATTTTCATGAGGAGTACGTGGCCGAGGAGTGC 669
QY 830 TCGGGGTGACTGTCAGAGTATAGGAGTACCACTACCTTCTCTTACCAAGAGGCGAG 889
Db 670 TGAGGGTCTACTGTGAGAGTGACAGGAGTACCACTCTCTCTTCTTACCAAGAGGCGAG 729
QY 890 AGGACTACCATATGACAGCAGTGAAGAACAAAGTCTACAGTTCTTCTAGTCGATC 949
Db 730 AGGACTATCCCTACGACAGTGGAGAACAAATCAAGTTCTGCAAGTTCTTCTAGTCGATC 789
QY 950 AGTTTCTTACAAACAATATTCGTCGAGAGGAATTCATGCTGAAGGGGTGATACAGTATG 1009
Db 790 AGTTCTTCTACGACCAATATTCGTCGAGGAGTGTATGCTGAAGGGGTGATACAGTATG 849
QY 1010 ATGACCATGTAGGGTGTGTACAAACTTGGGGATTTGCTTGTCTGTGAGACATGTTTCA 1069
Db 850 ACGACCATGTAGGGTGTGTACAAACTTGGGGATTTGCTTGTCTGTGAGACATGTTTCA 909
QY 1070 CAGTATACATTTGAAATGTGTGAAGCCACCTCTTGAAGAGGTGCGAGAGGAGGAGTGC 1129
Db 910 CTGTGTACATTTAGAGTGTGTGAAGCCGCTCTTGAAGAGTGTGCGGAGGAGGAGTGC 969
QY 1130 AGTGTGAAGTCTGTAGCACACAAAGGTGCCTTGGTGTGACTGACTGTGTGCTGCTGAAATTC 1189
Db 970 AGTGTGAAGTCTGTGGCACACAAAGGTGCCTTGGTGTGACTGACTGTGTGCTGCTGAAATTC 1029
QY 1190 AAAAAAATAAACCATATATTTCACATGAACCTATTGGGATATGATAGAAGTCCGAGAGAAAT 1249
Db 1030 AAAAGAAATAAACCTTACGTTTCGACACGAACCTATTGGGTATGACAGAGCCGCTAGGAAAT 1089
QY 1250 ACTGCTTCTTGAACCGAAGACTCATATAAGAGAAGATACACAAAATGAAATGAAAGA 1309
Db 1090 ACTGCTTCTTGAACCGAAGACTCATATAAGAGAAGACACACAAATGAAACGAGAGA 1149

Qy	8149	GAGCTGAAGAGACGACCTGRAAAATTAAGAAGAAAAAAGAACCTGTGCGATTGGCTCAGGCC	8208
Dd	1097	GAGCTGAAGAAGACACTGAAAATGAACGAGAGAGGGAGATGCCCCAGGCGGTACAGGCC	1156
Qy	8209	ACAGCAGTAGTCGCACCCCTGCCCCOCAGTGACACAGTTCCTTCCAGCCOCTCAGCCOCT	8268
Dd	1157	AATGCTGCCTCAGTGCCACACCTCCGTGCCAGCCOCTGTGCCAGCOCTGTGCCAGCGCA	1216
Qy	8269	CCACTTCACCTCCCCTCCACCTGTGTGCAACACACAGCGCTCTCTGT-----	8317
Dd	1217	GCOCCTCCAGCTCCTCTCTGTTCTCGCCTCCCTCCACACACAGTCTGCCACCTGCAGGC	1276
Qy	8318	----CCACGCCACCCTTACCTGTGTCTCCAGAAGAGGAAGCGGAAGAGGAAAAAGAC	8373
Dd	1277	CACCCACAGCCCACTGCCTGTCACTTCCCAAGAAGGAAGCGGGAGGAAGAAGGA--	1335
Qy	8374	TCCAGCTCAAGTCCAAGAAAAAAGAAATGATCTCTACTACCTCAAAAGGAAACTAAGAAG	8433
Dd	1336	-----CTCTAAGTCCAAGAAGAAAGATGATCTCTACCACCTCTAAGGAGGCCAAGAAG	1390
Qy	8434	GACACAAGCTTTACTGTATCTCTATAAACGCCCTTATGATGAATCTAAATTTTTATATTGGC	8493
Dd	1391	GACACCAGGCTATATTGCACTTCGAACACACCGTACGATGAGTCCCAAATTTTATATTGGC	1450
Qy	8494	TGTGATCGGTGTCAGAAATTTGGTACCATGGCGCTGGTTGGGATCTTTGCAAAGTGAGGCA	8553
Dd	1451	TGTGATCGGTGTCAGAAATTTGGTACCACGGCGCTGTGTGGCATCTTTGCAAAGTGAGGCA	1510
Qy	8554	GAGCTAATTCATCAGTATGTCGTCCACAGTGCAGCTCAACACAGGATGCCATGACACATG	8613
Dd	1511	GATCTCAATGTCAGTATGTCGTCCACAGTGCAGCTCGACACAGGACGCCATGACACATG	1570
Qy	8614	CTCACGCCACTAAACAGAAAGGATTAATGAGGGTTTGAAGGGTGTCTCCCTTCTTACAG	8673
Dd	1571	CTCACACCCTGACAGAGAAAGATTATGAGGGCTTGAAGGGTGTCTCCCTTCTTACAG	1630
Qy	8674	GCCCATAAGATGGCTGGCCCTTTCCCTTGAACCACTAGACACCCCTAATGATGCACAGATTAT	8733
Dd	1631	GCCCATAAGATGGCTGGCCCTTTCCCTTGAACCGGTAGACCCCAATGATGACCCGGATTAT	1690
Qy	8734	TATGGTCTTATTAAAGAACCTATGGACCTTTGCCACCATGGAAGAAAGAGTACAAAAGCA	8793
Dd	1691	TACGGTGTATTAAAGAGCCAATGGACCTTTGCCACCATGGAAGAAAGATACAAAACGG	1750
Qy	8794	TATTATGAAAGCTGACGGAATTTGTGGCAGATATGACCAAAAATTTTGTGATACTGTCGT	8853
Dd	1751	TATTATGAAAGCTGACAGAGTTCTGTGGCAGATATGACCAAAAATTTTGTGATACTGTCGT	1810
Qy	8854	TACTACAACTCCAGTGTACTCCCAATTTTACCAGTGTGCAGAAGTTCGCAATCAATCTTTT	8913
Dd	1811	TATTACAACTCCCGTGTACACCCCTTTTACCAGTGTGCAGAAGTTCGTGAATCAATCTTTT	1870
Qy	8914	GTACAGAAATTTGAAGGCTTCAAGCTTAGCAGGTCTCATAAACAACAACCTGCAGCTTACA	8973
Dd	1871	GTACAGAACTAAAAGGATTTCAAGGCCAGCAGGTCTCATAAACAACAAGCTGCAATCTACA	1930
Qy	8974	GCTTCTTAAAGTTCAGCGGTGTTAACTTAACTAATAACACACAGCAAGAATCTGGTTGCTGA	9033
Dd	1931	GCTCCTTTAGACTCAGCGTCTGTGTACCTTAAGCTAGACACAGCAAGCTTGGCGCTCTGA	1990
Qy	9034	ACTATTTTAAATTAAGGAGCCAGATGTTTTTGTAGTCAAGCTATCTGACAAAGACTTGACCT	9093
Dd	1991	ACTA-TTTAACTTAAGCGCAGATATTTTCAGTFCAGGCTTTCCTGACAA-----GACCG	2044
Qy	9094	AAACTTCGTTTTTATTTGGTCATACAGTCCCAATTTATTTCTTTGGCCCAATTTTGTCCAACG	9153
Dd	2045	TAACCTCGTTCATATTGTTGCACACACTCCAGTGTGATTTCTTGGCCCAATTTTGTCCAACG	2104
Qy	9154	GACAAGAAAAAGCAAGTCAACACACCAATTTATCTTGTCAAGATCAGATGGTTTACTA	9213
Dd	2105	GACAAGAAAAAGCAAGTCAACGGCACCGTGTCTTGTGCAGAGCAATTTGGCTTTACTA	2164

Qy 9214 TTGTGGCAGAAACGCGAGAAAACCTTTGTTTTATTTGAATAAAAAAAGAAAAAGA 9264
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2165 TTTGTGGCAGAAAGC-AGGAAACTTTTGTATTATGGAAAAAAAAAAAAAAAAA 2214

RESULT 15
AC109873 223387 bp DNA linear HTG 11-OCT-2002
LOCUS Rattus norvegicus clone CH230-303015, *** SEQUENCING IN PROGRESS
DEFINITION ***, 2 unordered pieces.
ACCION873 AC109873 4 GI:23605608
VERSION HTG; HTGS_PHASEI; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rats.

REFERENCE
AUTHORS 1 (bases 1 to 223387)
Muzny,D.Marle., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
Anylabechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Bliswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davilla,M.B., Davis,C., Davy-Carroil,L., De Anda,C., Dederich,D.,
Delgado,O., Dengen-Rocha,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Duncan-Schoja,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,W.,
Gebregorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Gusvava,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.I., Hodgson,A., Hoques,M.,
Holkins,B., Howells,S., Huliyk,S., Hume,J., Idlebirt,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Joliveta,A.,
Karpthy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levam,J., Lewis,L., Li-Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Manung,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris.S.,
Nsakailemeh,O., Okwuonu,G., Olarnpunagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzao,M., Quirioz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rokey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shivartsbeyn,A., Sisson,I., Sittler,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Soresle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabot,P., Taylor.C.,
Taylor.T., Thomas.N., Thomas.S., Tingey,A., Trejos,Z., Usmani.K.,
Valas.R., Vera.V., Villalana,D., Waldron,L., Walker.B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White.F.,
Williams.G., Willison,R., Wlaczky.R., Wooden.H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu.F., Zhang,J., Zhou,J., Zhou.X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 223387)
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
FILE

JOURNAL

Submitted (08-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 223387)
 Rat Genome Sequencing Consortium.
 Direct Submission

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Oct 9, 2002 this sequence version replaced gi:21738143.

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GOIO

Center clone name: CH230-303015

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 191060 bases at least Q40

Consensus quality: 195253 bases at least Q30

Consensus quality: 198285 bases at least Q20

Estimated insert size: 194499; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved

* 1 222161: contig of 222161 bp in length

* 222162 222261: gap of unknown length

* 222262 223387: contig of 1126 bp in length.

FEATURES

source

Location/Qualifiers

1. .223387

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-303015"

1. .1317

/note="wgs_end_extension"

clone_end:T7

/note="clone_boundary"

clone_end:T7

site:MboI

end_sequence:RXAKI92TJ"

4700. .5756

/note="wgs_contig"

29062. .30112

/note="wgs_contig"

43664. .45232

/note="wgs_contig"

101380. .102634

/note="wgs_contig"

misc_feature

195324. .197239

/note="wgs_contig"

misc_feature

complement(215803. .216669)

/note="clone_boundary"

clone_end:Sp6

site:MboI

end_sequence:RXAKI92TV"

220191. .222161

/note="wgs_end_extension"

clone_end:Sp6"

BASE COUNT 52000 a 44167 c 46629 g 57824 t 22767 others

ORIGIN

Query Match 14.7%; Score 1454.4; DB 2; Length 223387;
 Best Local Similarity 76.1%; Pred. No. 5.9e-223;
 Matches 1833; Conservative 0; Mismatches 566; Indels 9; Gaps 3;

QY 3210 AGATGTGAAGGAGCTCTTAGATCTGTGACAGTGTAAACCTTCAAGGAGCAACCAATGGA 3269
 Db 46425 AGATGTAAAGAACTTTTAGATCTGTGATCTTCAAGGAGCAACCAATGGA 46484
 QY 3270 AGTAGACGATGACATGAAACAGAGTCACATGTAAATTGTGAGAGAGTTCTCAAGTAGA 3329
 Db 46485 GATCGATGACAGTGTGAAACAGAGTCACATATGACAGTGTGAGAGAGTGCACAGGTAGA 46544
 QY 3330 TGTGGTCAATGTAGTGAGGGTTTTCATCTTAAGGACTAGTTACAAAAAGAAACAAATC 3389
 Db 46545 TGTGGTCAATGTAGTGAGGGTTTTCATCTTAAGGACTAGTTACAAAAAGAAACAAATC 46604
 QY 3390 ATCCAACTAGATGGACTTCTTGAAGGAGAAATTAACAGTTTACACTGGAAGAAACAA 3449
 Db 46605 CTCTAAGCTGGACGGCTTCTGGAACGGAGATTAGACAAATTTACACTGGAAGAAACAA 46664
 QY 3450 GCACACTCGAAAAAATCAAGTTGGAGGTGGAATTAAGGGTATAGGAAAGACTTCTACAAA 3509
 Db 46665 GCCTCTGAGAACCTTAAAGCTGGAGAGTGGGGTGAAGGGGAGGCAAGCCCCCAGG 46724
 QY 3510 TTCTTCAAAAAATCTCTCTGAATCACCAGTAAATTAACGAAAGCAAGAAAGGGTGTCAAG 3569
 Db 46725 TGCTGTAAAAAGTCTCTCTGAGCTCGGGTATCCAAAAAGCAGACGAAAGGACCAGG 46784
 QY 3570 TGACTCGATGACAGACAGACAGAGCCCAATATGCAATATGATCAACCTGAGGACTTGTAT 3629
 Db 46785 TGACTTGTCTTAGACAGAGCAGAGGTCCACTCCAGTCCAGCTGGCAGTCCGACTTTGGG 46844
 QY 3630 TCAGGGATGTTCAAAAGTCAATCTCAGTCTTTAGAAATCAGTGTATCCTAGTATACACAC 3689
 Db 46845 GCTGGGGGCTCACAGAGTATCCCTTGGTTCCTTGGCATCAGTCTCTCTAGTCTCAGCAC 46904
 QY 3690 AAACAACTTTATCCAAAAGATCGAGTGTGTAGATGTCTCCATTCGGAGCCCAAGAAC 3749
 Db 46905 GGGCAAACTGTATCCAAAAGTCAAGTATTAGATGTCTCTCCATTCAGAGTCCGGGGCC 46964
 QY 3750 AAAATGTCGAAACAAATTCATTTGAAATGACATGAGAAAGTCTCTGACCTTGC 3809
 Db 46965 AGCTGTACGGGGCAGAACTCTCTTGAAGCCAGCTAGATGTCAGGATCTCTGAAGGTAC 47024
 QY 3810 CAGTAGAGGCCAGGAACCCACTAAGAGTAAACCAAGGAAATGATTTTTTCATCGATGA 3869
 Db 47025 CAGTAAGGGACTGGAACCTAAGTCAGACAAAAACAAAGTACTGACTCTTTTACTGATGA 47084
 QY 3870 CTCCTAACTAGCCAGTGCAGATGATATTTGGTACTTTTGTATCTGTGAAGCAAAAAACCGCT 3929
 Db 47085 CTCCTAAACCAACCAAGTGCAGATGAGGTAGCACTCTGTCTGTAAAGCAAGAAGCGCT 47144
 QY 3930 CATACAGGAGGAAAGTGCACACCAATTTT---TTCTTCTTCCAGAGTCTTTTACATTCATC 3986
 Db 47145 CCCACAGGAGGATTTGAGCACTGCTGTGTCTATCTTCTTCCAGAGTACTTTAGCTGCACC 47204
 QY 3987 AGTGCCTTAAAGTACCAATGACAGAGATGCCACACCTCTCTGTCAAGAGCAATGACTTTGA 4046
 Db 47205 AGTCCCTTAAAGTCCCGCTGACAGAGATGCCCGGGCTTTTCCCGAAAGCAATGACTTTGA 47264

```
QY 4047 AGGAAACTGGGATGTGACTCTGAATCTAATAGACACTTTGGAAATAGTTCTGTATACCGT 4106
Db 47265 TGGAGACTAGGTGCTGACTCTGAATATATAGCACTTTGGAAATAACTCTGTGCTACTAT 47324
QY 4107 GTCTATTTCAGATAGCAGTGAAGATATGATTGTTTCAGAAATAGCAATGAAACATTTTC 4166
Db 47325 GTATATCGGGACACGACGTGAAGAGATATGTTTGTTCAGAAATAGCAATGACACACCTTC 47384
QY 4167 TGAACAGTTTCAGAACTCGAGAAACAGATGTTGAAGTCTTGAGGCGGTTAAAGTGTGAGTT 4226
Db 47385 TAAACGGTTCATACTCCAGACCAAGCATGGAAGTACGAGCGCGACAAATGTCTGATT 47444
QY 4227 GGTTCCTGGTGAAGTCCACTGGAACACTGTGAGGACAGGCTGCCGTCAAGGGACTGAAGC 4286
Db 47445 AGTTTC---TAAGTCCCACTGAAATTTGTGACACAGACTACAGGGTAAGGTAACCTGAAGC 47501
QY 4287 AAATGGTAAACCAAGTCCAGCAGAGAAATTAAGAGAGAGACCAAGTTAATRAATGTAG 4346
Db 47502 AAATGGTAAAGAACTGGGCCAGCATCAAAAGCCAGAGAGAGACGTTAAACAGATGTGC 47561
QY 4347 TGATCAATAAAGCTAAAAATAACCACTGACAAAAAGAAATTAATGAAAATCGAGAGTCTGA 4406
Db 47562 CGACCAAGTTAGCTTAAGACACAGCGCTGACAGAAAGAACAGTGAAGCCGAGAGTCTGA 47621
QY 4407 AAAGAAAGGACAGAGCAAGTACATTTCAATTAATGGAAGAAAGATAATAAACCCAAAT 4466
Db 47622 AAAGAGAGACAGAAAGCAATAAATTTTCAATTAATGGAAGAAAGTAAAGCTAAGGG 47681
QY 4467 ATATTGAAAGGTGAATCTTGAAGAAATTTCTGAGAGTAGAGTAGTGAAGTGTGAATGT 4526
Db 47682 GTATCTGAAGGGACCGGCATGAAGGATGTCTGTGATGGTAAAGTCATGAGTGGTGCCTG 47741
QY 4527 TGAACCAAGGTTAATATAATAAATAAATCCCTGAGAAATGATATTAATCAATTCATTGAC 4586
Db 47742 TGAACCAAGGTCATATATCAATAAAGTCAATCCCTGGAA--ATATCAAGTCACCTGGC 47798
QY 4587 TGTAAAGAAATCTGTATAAGGCCATTCATTAATGGTGATCTCATATGGAAGATTTTAA 4646
Db 47799 TGGGAAGGAATCTGTGTGAAGCCGTTTGTAAAGGGACATCATATGGAAGAGTCTCAG 47858
QY 4647 TGAAGAAACAGCTCCGAAACAAATCGATTGCTGAGTCTTCAGATGCTGAAGGTAA 4706
Db 47859 TGAACAAACACTTCAGAAACCACTCGTATTCTGAGTCTTCAGATGCCAAAGGTAA 47918
QY 4707 CTACGAGATAGCCTTGAGACCCCTGCCATCAACCAAGAGTCTGACAGTACACAGACGAC 4766
Db 47919 TCACCAAGATGGCTCCATACCTACCATCCACCAAGACCTGCGAGTACCAGGTAAT 47978
QY 4767 CACACCCCTCAGATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAATATGGAAT 4826
Db 47979 CATACCCCCAGCCCTTGTCCAGATCGGAACCTCACTTAATCAGGTAGAAATATGGAAC 48038
QY 4827 AGAAACCTTCAGAGTTAAGAAAGTTACTTCATCACTATTACTTCTGAAGAGGAATCTAA 4886
Db 48039 TGAAGGTCAGAGGTTAAGAAAGTTACCCGCTCTCCCATTTACTGAGAGAGGACTCTAA 48098
QY 4887 TCTCAGTAATGACTTTATTGATGAAATGGTCTGCCCATCAACAAAAATGAAATGTCAA 4946
Db 48099 CCTCGTTAAGAGCTGATGATGAAACGGTCTACCTTCCAGCAAGATGAGATGTCAA 48158
QY 4947 TGGAGAAATCTAAAGAAAAACCGTCATCAGAGAGTACACAGATGACCTCCACAGTGGC 5006
Db 48159 TGGAGAAATCCCAAGAAAGAGCGGTCAATCAGAAAGTCCACCAATGACATCCACTGTTGC 48218
QY 5007 CACAGAAATCAAAACTGTGATCAGGTAGAAAAAGGCGATAAGCAAACTGTGGTTTCTTC 5066
Db 48219 CACAGAAATCAAAACTGTCAATTAAGGTAGCAAAAGGGGACAAAGCAACTGTGGTTTCTTC 48278
QY 5067 CACAGAAATTTGTCAAAATCCACTGTGCACAAACCACTACAAAGTGAACCAAGCTTTC 5126
Db 48279 CACAGAAATCTGTGCCAGATCCACTGTGCACAACTACCACCACTGTGTGACAAAGCTCTC 48338
QY 5127 CACACCCCTCCAGCGCGCAGTGTGGACATCATCTCTGTAAGGAGAGACGCAAAACCGT 5186
```

Search completed: September 23, 2003, 15:44:36
Job time : 23092 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:12:14 ; Search time 1522 Seconds
(without alignments)
17496.686 Million cell updates/sec

Title: US-09-698-295-2
Perfect score: 9865
Sequence: 1 ggcacggtgagtggtgccc.....acaaaaaaaaaaaaaaaaa 9865

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/cgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/cgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/cgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/cgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/cgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/cgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9865	100.0	9865	21	Human transcriptio
2	9067	91.9	9700	21	Human transcriptio
3	7255.2	73.5	8280	25	Human nucleic acid
4	2565.6	26.0	2673	24	Human cDNA differe
5	1719.4	17.4	2057	22	Human cDNA sequenc
6	1240.2	12.6	1540	22	Human diagnostic a
7	814.6	8.3	1032	20	Human secreted pro
8	771.4	7.8	841	22	Human cDNA clone (

9	735	7.5	735	24	ABK47549	DNA encoding SCAM/
10	673.8	6.8	1408	21	AAC99129	Human pancreatic c
11	627.4	6.4	750	22	AAI94524	Human neuroblastom
12	574	5.8	574	25	ACA57390	Human adipocyte Se
13	574	5.8	2123	22	AAH17323	Human cDNA sequenc
14	538.2	5.5	556	22	AAH09968	Human cDNA clone (
15	413	4.2	763	20	AAZ17218	Lung small cell ca
16	359	3.6	574	24	AA561658	Human gene express
17	355.6	3.6	368	21	AAC07933	Human secreted pro
18	348.4	3.5	15037	22	AAI99283	Human excretory re
19	348.4	3.5	15037	22	AAI63633	Human kidney relat
20	346.2	3.5	347	21	AC30831	Human secreted pro
21	316.4	3.2	341	24	ABK45859	cDNA encoding colo
22	303	3.1	731	22	AAI94525	Human neuroblastom
23	263	2.7	265	14	AAQ59932	Human brain expres
24	244.8	2.5	2928	23	ABL23837	Drosophila melanog
25	244.8	2.5	5380	23	ABL23836	Drosophila melanog
26	222.8	2.3	865	23	AA583023	DNA encoding novel
27	220.8	2.2	247	21	AA45848	Human secreted exp
28	214	2.2	214	21	AAC03141	Human secreted pro
29	181.4	1.8	405	24	ABS69806	Novel murine polyn
30	177.4	1.8	2666	24	ABK35890	cDNA sequence #281
31	177.4	1.8	2671	21	AAZ34559	Human secreted pro
32	177.4	1.8	4845	22	AA535982	Human cardiovascular
33	177.4	1.8	4646	22	AA535981	Human dithp transc
34	176.4	1.8	411	25	ACC46298	Human colon cancer
35	175	1.8	452	22	AAH36921	Human gene signatu
36	162.8	1.7	167	16	AAT19304	Human pancreatic c
37	157.4	1.6	311	24	ABV95834	Human secreted pro
38	120.4	1.2	242	21	AA09489	Oligonucleotide fo
39	101.4	1.0	618	24	ABQ47666	Oligonucleotide fo
40	101.4	1.0	618	24	ABQ47667	Oligonucleotide fo
41	95.6	1.0	2465	11	AAQ03665	Sequence homology
42	91.2	0.9	589	24	ABQ50266	Oligonucleotide fo
43	91.2	0.9	589	24	ABQ50267	Oligonucleotide fo
44	90.8	0.9	157	24	ABZ08622	Human leukocyte de
45	90.2	0.9	360	23	ABL30495	Drosophila melanog

ALIGNMENTS

RESULT 1
AAZ39032
ID AAZ39032 standard; cDNA; 9865 BP.
XX
AC AAZ39032;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human transcriptional regulatory factor TCoal gene SEQ ID NO:2.
XX
KW Human; transcriptional regulatory factor; TCoal; BLAST detection;
KW bromo-domain; cell proliferation; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 259..8982
FT /*tag= a
FT /product= "transcriptional regulatory factor"

XX
PN WO9957143-A1.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-JP02340.
XX
PR 30-APR-1999; 98JP-0137631.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Jones MH;

XX WPI: 2000-052940/04.
DR P-PSDB: AAY57452.
XX
XX Transcriptional regulatory factor containing a bromo domain and gene
XX TCoA1 encoding it -
XX
XX Claim 4; Page 69-98; 154pp; Japanese.
XX
XX The present sequence encodes a human transcriptional regulatory factor
XX containing a bromo domain. The factor interacts with proteins involved
XX in the chromatin-mediated transcription regulatory mechanism. It binds
XX to hSNF2H, hSNF2L and NCoA-62/Skip. It can be used for screening
XX compounds binding to it and acting as agonists or antagonists, which
XX are potentially useful for the treatment and prevention of cancer and
XX other cell proliferation disorders.
XX
XX Sequence 9865 BP; 3338 A; 2121 C; 2218 G; 2188 T; 0 other;
Query Match 100.0%; Score 9865; DB 21; Length 9865;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9865; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCAGGCTGAGGTGGCGCCCAAGACGCGGCTGAGCTCGCCAGGGGGGCGCAGCTAGC 60
DB 1 GCCCAGGCTGAGGTGGCGCCCAAGACGCGGCTGAGCTCGCCAGGGGGGCGCAGCTAGC 60
QY 61 CGGAGGAAGCCGCCCGCGCCGCCCGCGCCCGCCCGCCCGCCCGCGCGCGCGCGGG 120
DB 61 CGGAGGAAGCCGCCCGCGCCGCCCGCGCCCGCCCGCCCGCCCGCGCGCGCGGGG 120
QY 121 GGGCAGGAGCGCGGGCGGCACACAGCGGGGGCGGGCGGGCGGGCGGCCACCTGTCCCGG 180
DB 121 GGGCAGGAGCGCGGGCGGCACACAGCGGGGGCGGGCGGGCGGGCGGGCGGCCACCTGTCCCGG 180
QY 181 ACCACCGCGCCCGAGAGGCGGTCAACAAAGTGTGTACGATGACACAGAGAGCGGAGCGG 240
DB 181 ACCACCGCGCCCGAGAGGCGGTCAACAAAGTGTGTACGATGACACAGAGAGCGGAGCGG 240
QY 241 GTGGAGGAGAGGAGGACATGCTCCGAGGAGGAGGAGGAGGAGGAGCGGCGCGCGAG 300
DB 241 GTGGAGGAGAGGAGGAGGACATGCTCCGAGGAGGAGGAGGAGGAGGAGGAGCGGCGAG 300
QY 301 GAGACCCAGGATTTCTGAGGACGACGAGGAGGATGAGATGGAAGAGGAGGAGGATGCTCC 360
DB 301 GAGACCCAGGATTTCTGAGGACGACGAGGAGGATGAGATGGAAGAGGAGGAGGATGCTCC 360
QY 361 GATTATCCGGAGGAGATGGAAGACGACGACGACGACGACGACGACGACGACGACGACG 420
DB 361 GATTATCCGGAGGAGATGGAAGACGACGACGACGACGACGACGACGACGACGACGACG 420
QY 421 TTCAGGAGCCATGATGACCTACAGCAGCAGCTCCAGGTAGGCGAAACCAAGAGTACATCGG 480
DB 421 TTCAGGAGCCATGATGACCTACAGCAGCAGCTCCAGGTAGGCGAAACCAAGAGTACATCGG 480
QY 481 CUTCCTTCTCCTATATTGGAAGAAAGACATCCCGCCCTTGAAATTTCCCAAGTCTCT 540
DB 481 CUTCCTTCTCCTATATTGGAAGAAAGACATCCCGCCCTTGAAATTTCCCAAGTCTCT 540
QY 541 GAGGATTTAATGGTCCCTAATGAGCATATATGATGTCATGCCATTTACGAGTACTG 600
DB 541 GAGGATTTAATGGTCCCTAATGAGCATATATGATGTCATGCCATTTACGAGTACTG 600
QY 601 CGGAATTTGGGACCTGTTTGGAGATTATCTCTTTTCGCTTTGAGGACTTTTGTGCAGCT 660
DB 601 CGGAATTTGGGACCTGTTTGGAGATTATCTCTTTTCGCTTTGAGGACTTTTGTGCAGCT 660
QY 661 CTGGTGAGCCAAAGAGAGTGTCACATCATGCGCAGAGATGCTGTTGCTTTGGAAGCA 720
DB 661 CTGGTGAGCCAAAGAGAGTGTCACATCATGCGCAGAGATGCTGTTGCTTTGGAAGCA 720
QY 721 GTTCTCGGTGAGAGACACTTCCCAATCTACTCTTTGGACCTGCTGATCTGAAAGATAGC 780
DB 721 GTTCTCGGTGAGAGACACTTCCCAATCTACTCTTTGGACCTGCTGATCTGAAAGATAGC 780

DB 721 GTTCTCGGTGAGAGACACTTCCCAATCTACTCTTTGGACCTGCTGATCTGAAAGATAGC 780
QY 781 GTTAATTCACACTGTATTTTCATAGATGGGATGACGTGCCAGAGAGTGTCTGGGGGTGAC 840
DB 781 GTTAATTCACACTGTATTTTCATAGATGGGATGACGTGCCAGAGAGTGTCTGGGGGTGAC 840
QY 841 TGTGAGAGTGATAAGAGGTACCATCAGTTCTTCTTACCAAGAGCGAGAGACTACCCA 900
DB 841 TGTGAGAGTGATAAGAGGTACCATCAGTTCTTCTTACCAAGAGCGAGAGACTACCCA 900
QY 901 TATGGACCAGTAGAGAACAAAGATCAAAAGTTCTACAGTTTCTAGTGCAGTCTTCTTACA 960
DB 901 TATGGACCAGTAGAGAACAAAGATCAAAAGTTCTACAGTTTCTAGTGCAGTCTTCTTACA 960
QY 961 ACAAAATATTGCTCGAGAGGAATTGATGCTCTGAAGGGGTGATACAGTATGATGATGAT 1020
DB 961 ACAAAATATTGCTCGAGAGGAATTGATGCTCTGAAGGGGTGATACAGTATGATGATGAT 1020
QY 1021 AGGGTTTGTACAAAATCTGGGGATTTGCTTGTGTGACACATGTTTCAGCAGTATACCAT 1080
DB 1021 AGGGTTTGTACAAAATCTGGGGATTTGCTTGTGTGACACATGTTTCAGCAGTATACCAT 1080
QY 1081 TTTGGAATGTGTGAAGCCACTCTTTGAGGAGGTGCCAGAGAGAGAGTGGCAGTGTGAAGTC 1140
DB 1081 TTTGGAATGTGTGAAGCCACTCTTTGAGGAGGTGCCAGAGAGAGAGTGGCAGTGTGAAGTC 1140
QY 1141 TGTGTAGCACACAAAGTGCCTGCTGTGACTGACTGTGTTGCTGAAATCCAAAATAATAA 1200
DB 1141 TGTGTAGCACACAAAGTGCCTGCTGTGACTGACTGTGTTGCTGAAATCCAAAATAATAA 1200
QY 1201 CCATATATTGCGACATGAACTTATGATATGATAGAGTGGAGGAAATACTGTTCTTGTG 1260
DB 1201 CCATATATTGCGACATGAACTTATGATATGATAGAGTGGAGGAAATACTGTTCTTGTG 1260
QY 1261 AACCGAAGACTCATAATAGAAAGATACAGAAATGAAATGAAATGAAATGAAATGAAAT 1320
DB 1261 AACCGAAGACTCATAATAGAAAGATACAGAAATGAAATGAAATGAAATGAAATGAAAT 1320
QY 1321 TACAGCACAAAGGTCCAACTTTCAGAGATTAATTTGACTGTCTAGACAAAGATTTATGCG 1380
DB 1321 TACAGCACAAAGGTCCAACTTTCAGAGATTAATTTGACTGTCTAGACAAAGATTTATGCG 1380
QY 1381 GCAGAACTCTGCAAAATTTCTAGAAAGAAATCGGTGAAGAAATCCACCCACACATGGACA 1440
DB 1381 GCAGAACTCTGCAAAATTTCTAGAAAGAAATCGGTGAAGAAATCCACCCACACATGGACA 1440
QY 1441 ACTGAAGACCTGACCAATTAAGGCTCGGGGCGAGTAACAAATCCTTTCTGGCGGCGAGCT 1500
DB 1441 ACTGAAGACCTGACCAATTAAGGCTCGGGGCGAGTAACAAATCCTTTCTGGCGGCGAGCT 1500
QY 1501 GAAGAAATTTTGGATCCATTAAGAGCCAAAGAGGAGACATTTGATATGTTTAAAGGCCA 1560
DB 1501 GAAGAAATTTTGGATCCATTAAGAGCCAAAGAGGAGACATTTGATATGTTTAAAGGCCA 1560
QY 1561 GAAGAAACAGAAAAGACAAAGATGAGACTGAGAACTGACTCTAAAGATGCTGAGAAAAC 1620
DB 1561 GAAGAAACAGAAAAGACAAAGATGAGACTGAGAACTGACTCTAAAGATGCTGAGAAAAC 1620
QY 1621 AGAAGAAATTTGAAGACCTGCTTGAAGAGAGTGCAGCAGCAAAAACACAGATGAT 1680
DB 1621 AGAAGAAATTTGAAGACCTGCTTGAAGAGAGTGCAGCAGCAAAAACACAGATGAT 1680
QY 1681 GACCTTGAGCAAGGAAATCTGAGAGGCCAAACAGAGTTGGGGATTAAGGTAACTCTGTG 1740
DB 1681 GACCTTGAGCAAGGAAATCTGAGAGGCCAAACAGAGTTGGGGATTAAGGTAACTCTGTG 1740
QY 1741 TCAGCAAAATCTTGGCGACACACAAACAAATGCAACTTTCAGAAAGAGACTAGTCCCTCTG 1800
DB 1741 TCAGCAAAATCTTGGCGACACACAAACAAATGCAACTTTCAGAAAGAGACTAGTCCCTCTG 1800
QY 1801 GGGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCCCGCATAGCAGCAACATGGCAGAGAAG 1860
DB 1801 GGGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCCCGCATAGCAGCAACATGGCAGAGAAG 1860

QY	1861	AAGGTGGCATCTGAGCTCCCGAGGATGTCAGAGAACCTTAACAAGACATGTGAGAGC	1920	2941	TGGAGCTGATAGTAAAACTCATGTTTATAGTGTTCCTTAAATTTGCCAGGCAATACT	3000
Db	1861		1920	2941		3000
QY	1921	AGGTGGCATCTGAGCTCCCGAGGATGTCAGAGAACCTTAACAAGACATGTGAGAGC	1920	3001	AATGTGAATTACAGAAAGTCGTAGAGAACCAAAAATAAATATGATGAAATATGGAT	3060
Db	1921		1920	3001		3060
QY	1981	AGTGAACCTTCTTCCAGAGTGAATCTGTAAGGCAGCTGATGATCCTGAAATGGA	2040	3061	GAGTCAGATATAAGAAAAATGTTTACGAAGTCCAAAAATAAATAAGAGCCTGATCT	3120
Db	1981		2040	3061		3120
QY	2041	GAAAGAGATCTATACACCTCTCTATTCAGGAGAGATAGTAGGTTGATTTACATCG	2100	3121	GAAAAAGATGAGTAAAAAGTTTACGAAGTCCAAAAAGAGCAGACAAAATAAGATGGAT	3180
Db	2041		2100	3121		3180
QY	2101	GAGAGTCCACGGGGAGCTAAGTGAATCTCTGGAGCTGGAAGAGGACATCTGGCTCA	2160	3181	ATCTCAAGATTTACTGAGAGAGGACCAAGATGTAAGGAGCTCTTAGATTCTGCAGT	3240
Db	2101		2160	3181		3240
QY	2161	ACTGAATCATCACCAGATTGCGGAATCCAGATAGCAAACTTAGTCAGTGAAGGCCAG	2220	3241	GATAAACCTTCAAGGAGAACCAATGGAAGTAGACGATGACATGAAAAACAGAGTCACAT	3300
Db	2161		2220	3241		3300
QY	2221	CAGGTGGCAGCGCTGCACATGAAGCAATAAATATTTAAGGAGGGCAAGAGGTACTG	2280	3301	GTAAATTTGTGAGGAGATTCTCAAGTAGATGGTCAATGTTAGTGAGGGTTTTCATCTA	3360
Db	2221		2280	3301		3360
QY	2281	GTAGTTAACTCTCAAGGAGAAATTTACGGTTGAGCACCACCAAGAGGTATCATGAA	2340	3361	AGGACTAGTTACAAAAGAAAAACAATCATCCAACTAGATGGAGCTCTTGAAGGAGA	3420
Db	2281		2340	3361		3420
QY	2341	GGAAATATCAACAATATTTAAATTTGGTCAAGAGGAGGATATCGCTCTACACAAAT	2400	3421	ATTAAACAGTTTACACTGGAAGAAAAACAGCGACTCGAAAAATCAAGTTGGAGGTGGA	3480
Db	2341		2400	3421		3480
QY	2401	CAATACTCCACAAATTCATTTGCTTTGAATAAGCACCAGCACAGAGAACCATGATAAG	2460	3481	ATTAGGGTATAGGAAAGACTTCTACAAATTTCTTCAAAAATCTCTCTGAATCACCAGTA	3540
Db	2401		2460	3481		3540
QY	2461	AGAAGCATCTTGGCATAGTCTGCTGCTGCTCCAGCAGGAGAGTTCAATGGACGGT	2520	3541	ATAACGAAAGCAAAAGAGGGTGTCTAGAGTCTGAGTCCGATGAGCAACAACAGAGCCCAAT	3600
Db	2461		2520	3541		3600
QY	2521	TCTGTCCATGGTCCAAAGTCTTACCATATCTACTCTGAGACTGACTATCACCCAATTA	2580	3601	GCAATAATGATCAACCTGAGGACTTTGATTGAGGGATGTTTCAAAAAGTGATTCCTCAGTT	3660
Db	2521		2580	3601		3660
QY	2581	GAAACACATCCCTTCATCTTTTTCATCCCACTGGGCAATCATAGGCAATTTGG	2640	3661	CTTAGAATGAGTGTCTAGTCTATACCAACAACAACTTTATCCAAAAGATCGAGTGA	3720
Db	2581		2640	3661		3720
QY	2641	ATCAAGGCAGTTCAGATGTAGCAAAACCCAGAGAAATTTGCATTTGACCATTTTG	2700	3721	GATGATGCTCTCCATTCGGAGCCCAGAAAATAAATGTCGGAACAAAATTTCCATTGAAAT	3780
Db	2641		2700	3721		3780
QY	2701	GAGTGTGAGTTAAACCCAGTTGCTGCTTACCAATATGGCGAGAAATTTTAGGACATAC	2760	3781	GACATAGAGAAAAAGTCTCTGACCTTGCAGTAGAGGCCAGGAACCCCTTAAGAGTAA	3840
Db	2701		2760	3781		3840
QY	2761	AGGTTTACCCGGATGACATCAATGAAAGAGAGAAAAAGGAGAAAGTCAAAAAAAGAG	2820	3841	ACCAAAGGAAATGATTTTTTTCATCGATGACTCTAAACTAGCCAGTCAGATGATTTGGT	3900
Db	2761		2820	3841		3900
QY	2821	RAGAACAGGAGAGAGAAACGATGACGAGGACATGCGGTAAATACACATTTTCA	2880	3901	ACTTTGATCTGTAAGAACAAAAACCGCTCATACAGGAGGAAAGTGACACCATTTGTTCT	3960
Db	2821		2880	3901		3960
QY	2881	GTAAAGCATCAGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTACAGGATATGGTGT	2940	3961	TCCTTCAAGAGTCTTTTACATTCATCAGTGCCTTAAAGTACCAATGACAGAGATGCCACA	4020
Db	2881		2940	3961		4020
				4021	CCTCTGTCAAGAGCAATGGACTTTTGAAGGAAAACTGGGATGTGACTCTCTGAATCTAATAGC	4080

CC other cell proliferation disorders.

SQ Sequence 9700 BP; 3235 A; 2127 C; 2199 G; 2139 T; 0 other;

Query Match 91.9%; Score 9067; DB 21; Length 9700;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 9480; Conservative 0; Mismatches 5; Indels 380; Gaps 3;

QY	2	GCCAGGCTGAGGTGGCGCCCAAGACGGCGCTGAGCTCGCCAGGGGGGCGCAGTAGTACC	61
DB	215	GCCAGGCTGAGGTGGCGCCCAAGACGGCGCTGAGCTCGCCAGGGTGGCGCAGTAGTACC	274
QY	62	GGAGGAAGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG	121
DB	275	GGAGGAAGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG	334
QY	122	GCCGAGGAGCGGGGGCGGCACGACGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	181
DB	335	GCCGAGGAGCGGGGGGGCGGCACGACGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	394
QY	182	CCACCGCGCCCGGAGGGCCGTCAACAAAGTGGTACGATGACCCACGAGCGAGGGCGG	241
DB	395	CCACCGCGCCCGGAGGGCCGTCAACAAAGTGGTACGATGACCCACGAGCGAGGGCGG	454
QY	242	TGGAGGAAGAGGAGACATGTCTCCGAGGAGGAGGAGGAGGAGGAGCGCGCCGAGG	301
DB	455	TGGAGGAAGAGGAGGAGACATGTCTCCGAGGAGGAGGAGGAGGAGGAGCGCGCCGAGG	514
QY	302	AGACCCAGGATCTTGAGGACGACGAGGAGGATGAGATGGAAGAGGAGGAGGAGGATGCTCG	361
DB	515	AGACCCAGGATCTTGAGGACGACGAGGAGGATGAGATGGAAGAGGAGGAGGAGGATGCTCG	574
QY	362	ATTATCCGAGGAGATGGAAGACGACGACGACGCGCCAGTTACTGCGGAAAGACGCT	421
DB	575	ATTATCCGAGGAGATGGAAGACGACGACGACGCGCCAGTTACTGCGGAAAGACGCT	634
QY	422	TCAGGAGCCATAGTACCTACAGCAGCAGCTCCAGGTAGCGGAAACCAAGATACATCGG	481
DB	635	TCAGGAGCCATAGTACCTACAGCAGCAGCTCCAGGTAGCGGAAACCAAGATACATCGG	694
QY	482	CTCGTTCCCTATATTGGAAGAAAAGACATCCCGCCCTTGAATTTCCCAAGTCCCTCG	541
DB	695	CTCGTTCCCTATATTGGAAGAAAAGACATCCCGCCCTTGAATTTCCCAAGTCCCTCG	754
QY	542	AGGATTTAATGGTGCCTAATGAGCATATAATGAATGTCATTTGCCATTTACGAGTACTGC	601
DB	755	AGGATTTAATGGTGCCTAATGAGCATATAATGAATGTCATTTGCCATTTACGAGTACTGC	814
QY	602	GGAATTTGGCAGCTGTTTGGATATATCTCTTTTCGCTTTGAGGACTTTTGTGAGCTC	661
DB	815	GGAATTTGGCAGCTGTTTGGATATATCTCTTTTCGCTTTGAGGACTTTTGTGAGCTC	874
QY	662	TGGTGAGCCAAAGAGCAGTGCACACTCATGGCAGAGATGCATGTTGCTTTTGAAGCAG	721
DB	875	TGGTGAGCCAAAGAGCAGTGCACACTCATGGCAGAGATGCATGTTGCTTTTGAAGCAG	934
QY	722	TTCTGCGTGAAGAGACACTTCCAACTACTACTTTTGGACCTGCTGATCTGAAAGATAGCG	781
DB	935	TTCTGCGTGAAGAGACACTTCCAACTACTACTTTTGGACCTGCTGATCTGAAAGATAGCG	994
QY	782	TTAATTCACACTGTTATTTGATAGATGGGATGAGTGGCCAGAGGTGTCGGGTGTACT	841
DB	995	TTAATTCACACTGTTATTTGATAGATGGGATGAGTGGCCAGAGGTGTCGGGTGTACT	1054
QY	842	GTGAGAGTGAAGAGTACCATCACGTTCTTCCCTTACCAGAGGCGAGGACTACCCAT	901
DB	1055	GTGAGAGTGAAGAGTACCATCACGTTCTTCCCTTACCAGAGGCGAGGACTACCCAT	1114
QY	902	ATGGACCAGTAGAGAACAAGATCAAGTTCTACAGTTTCTAGTCGATCAGTTCTTCAAA	961
DB	1115	ATGGACCAGTAGAGAACAAGATCAAGTTCTACAGTTTCTAGTCGATCAGTTCTTCAAA	1174
QY	962	CAATATTTGCTCGAGAGGAATTGATGCTGAAGGGGTGATACAGTATGATGACCATTTGTA	1021

DB	1175	CAATATTTGCTCGAGAGGAATTGATGCTGAAGGGGTGATACAGTATGATGACCATTTGTA	1234
QY	1022	GGTTTGTCAAACTTGGGATTTGCTTGTGTGAGACATGTTTGCAGCAGTATACCATTT	1081
DB	1235	GGTTTGTCAAACTTGGGATTTGCTTGTGTGAGACATGTTTGCAGCAGTATACCATTT	1294
QY	1082	TGGAATGTGAAGCCACTCTTTGAGGAGTGCCAGAGACGAGTGGCAGTGTGAAGTCT	1141
DB	1295	TGGAATGTGAAGCCACTCTTTGAGGAGTGCCAGAGACGAGTGGCAGTGTGAAGTCT	1354
QY	1142	GTGTAGCACACAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1201
DB	1355	GTGTAGCACACAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1414
QY	1202	CATATATTCGACATCAACCTTATTTGGATATGATAGAAGTGGAGGAAATACTGTTCTTGA	1261
DB	1415	CATATATTCGACATCAACCTTATTTGGATATGATAGAAGTGGAGGAAATACTGTTCTTGA	1474
QY	1262	ACCGAAGACTCATATAGAAGAATACAGAAAATGAAAATGAAAAGAAAATTTGGTATT	1321
DB	1475	ACCGAAGACTCATATAGAAGAATACAGAAAATGAAAATGAAAAGAAAATTTGGTATT	1534
QY	1322	ACAGCACAAAGGTCCAACTTGCAGAAATTAATTTGACTGCTAGACAAAAGATTTATGGGAAG	1381
DB	1535	ACAGCACAAAGGTCCAACTTGCAGAAATTAATTTGACTGCTAGACAAAAGATTTATGGGAAG	1594
QY	1382	CAGAACTCTGCAAAATTTCTAGAAGAAATCCGTGAAGAAATCCACCGACACATGGACATAA	1441
DB	1595	CAGAACTCTGCAAAATTTCTAGAAGAAATCCGTGAAGAAATCCACCGACACATGGACATAA	1654
QY	1442	CTGAAGACCTGACCAATTAAGGCTCGGGCAGTAAACAAATCCCTTCTGCGGCGAGTAATG	1501
DB	1655	CTGAAGACCTGACCAATTAAGGCTCGGGCAGTAAACAAATCCCTTCTGCGGCGAGTAATG	1714
QY	1502	AAGAAATTTTGAATCCATAGAGCCAAAGGGGAGACATTGATATGTTTAAAGCCCGAG	1561
DB	1715	AAGAAATTTTGAATCCATAGAGCCAAAGGGGAGACATTGATATGTTTAAAGCCCGAG	1774
QY	1562	AAGAAACAGAAAAGACAAAGATGAGACTGAGAATGACTCTAAAGATGCTGAGAAAACA	1621
DB	1775	AAGAAACAGAAAAGACAAAGATGAGACTGAGAATGACTCTAAAGATGCTGAGAAAACA	1834
QY	1622	GAGAAGATTTGAAGACAGTCCCTTTGAAAAGACAGTGAAGACAAAACACAGATGATG	1681
DB	1835	GAGAAGATTTGAAGACAGTCCCTTTGAAAAGACAGTGAAGACAAAACACAGATGATG	1894
QY	1682	ACCTTGACCAAGGAAAATCTGAGGAGCCAAACAGAGTTGGGGATTAAGGTACTCTGTCT	1741
DB	1895	ACCTTGAGCAAGGAAAATCTGAG-----	1917
QY	1742	CAGCAAACTTTGGCGACAAACACAACTTCAGAAAGAGACTAGTCCCTCTCTGAAG	1801
DB	1918	-----	1917
QY	1802	GGAGGAGCCCTGTGGGGTGTCTCTCAGAAAACCCCGATAGCAGCAACATGGCAGAGAAGA	1861
DB	1918	-----	1917
QY	1862	AGGTGGCATCTGAGCTCCCCCAGGATGTGCCNAGANAACCTTAACAGACATGTGAGAGCA	1921
DB	1918	-----	1917
QY	1922	GTAACACTAGTGTACCCTACCTCCATCCAGCCTAATCTGGAAGAAACAGTAACAGCAGCA	1981
DB	1918	-----	1917
QY	1982	GTGAACATAATTTCTCCAGAGTGAATCTGCTAAGGCGAGCTGATGATCCTGAAAATGGAG	2041
DB	1918	-----	1917
QY	2042	AAAGAGAACTCATACACCTGTCTCTATTTACAGGAAGAGATAGTAGTGTTCACATCGG	2101

1918	-----CTAGGTGATTTCAAAATGGG	1936	Qy	3182	TCTCAAAAGATTACTGAGAAGAGGACC	3241	Qy
2102	AGAAAGTCCACCGGGAGCTAAAGTGAATCTCCTGGAGCTGGAAAAAGGAGCATCTGCCTCAA	2161	Qy	3017	TCTCAAAAGATTACTGAGAAGAGGACC	3076	Db
1937	AGAAAGTCCAAACGGGGAGCTAAAGTGAATCTCCTGGAGCTGGAAAAAGGAGCATCTGCCTCAA	1996	Db	3242	ATAAACCCTCGAAGGAAGAACCAATGG	3301	Qy
2162	CTCGAATCATCAACAGATTGCGGAATCCAGATAGCAAACTTAGTCAGCTGGAAGAGCCAGC	2221	Qy	3077	ATAAACCCTCGAAGGAAGAACCAATGG	3136	Db
1997	CTCGAATCATCAACAGATTGCGGAATCCAGATAGCAAACTTAGTCAGCTGGAAGAGCCAGC	2056	Db	3302	TAAATTTGTCAGGAGAGTTCTCAAGTAG	3361	Qy
2222	AGGTGGCAGCCGCTGCACATGAAGCAAAATAAATATTTAAGGAGGCAAGAGGACTGG	2281	Qy	3137	TAAATTTGTCAGGAGAGTTCTCAAGTAG	3196	Db
2057	AGGTGGCAGCCGCTGCACATGAAGCAAAATAAATATTTAAGGAGGCAAGAGGACTGG	2116	Db	3362	GGACTAGTTTACAAAAAGAAAAACAAAT	3421	Qy
2282	TAGTTAACTCTCAAGCAGAAATTCACGTTGAGCAGCAAAAGGAAGTATCATGAAG	2341	Qy	3197	GGACTAGTTTACAAAAAGAAAAACAAAT	3256	Db
2117	TAGTTAACTCTCAAGCAGAAATTCACGTTGAGCAGCAAAAGGAAGTATCATGAAG	2176	Db	3422	TTAAACAGTTTACACTGGGAAGAAACAC	3481	Qy
2342	GAATATCAACAATTAATTTAAATTTGGGTCAAGAAGGGAAGTATCGCGTCTACCAATC	2401	Qy	3257	TTAAACAGTTTACACTGGGAAGAAACAC	3316	Db
2177	GAATATCAACAATTAATTTAAATTTGGGTCAAGAAGGGAAGTATCGCGTCTACCAATC	2236	Db	3482	TTAAGGGTATAGAAAGACTTCTACAAAT	3541	Qy
2402	AATACTCCACCAATTCATTTGCTTTGAATTAAGCACCACAGACAGAGAACCATGATAAGA	2461	Qy	3317	TTAAGGGTATAGAAAGACTTCTACAAAT	3376	Db
2237	AATACTCCACCAATTCATTTGCTTTGAATTAAGCACCACAGACAGAGAACCATGATAAGA	2296	Db	3542	TAACGAAACAAAGAGGGTGTCTAGAGTG	3601	Qy
2462	GAAGGCATCTTGACATAAGTTCTGTCTGACTCCAGCAGGAGAGTTCAAAATGGAACGGTT	2521	Qy	3377	TAACGAAACAAAGAGGGTGTCTAGAGTG	3436	Db
2297	GAAGGCATCTTGACATAAGTTCTGTCTGACTCCAGCAGGAGAGTTCAAAATGGAACGGTT	2356	Db	3602	CAAAATATCATCAACTGAGGACTTGTATTC	3661	Qy
2522	CTCTCCATGGGTCCAAAGTCTTACCATATCTACTCTGAGACTGACTATACCCCAATTAG	2581	Qy	3437	CAAAATATCATCAACTGAGGACTTGTATTC	3496	Db
2357	CTCTCCATGGGTCCAAAGTCTTACCATATCTACTCTGAGACTGACTATACCCCAATTAG	2416	Db	3662	TTAGATGAGTGATCCTAGTCTATACCA	3721	Qy
2582	AAACACATCCCTTCATCCTTTTTCATCCCACTGGGCATCACATAGGGCAAAATGGA	2641	Qy	3497	TTAGATGAGTGATCCTAGTCTATACCA	3556	Db
2417	AAACACATCCCTTCATCCTTTTTCATCCCACTGGGCATCACATAGGGCAAAATGGA	2476	Db	3722	ATGATGTCCTCAATTCGGAGCCAGAAAC	3781	Qy
2642	TCAAGCAGTTACAGATGTGTAGCAAAACCCAGAGAAATTTGCATTGGCTTTAGCCATTTGG	2701	Qy	3557	ATGATGTCCTCAATTCGGAGCCAGAAAC	3616	Db
2477	TCAAGCAGTTACAGATGTGTAGCAAAACCCAGAGAAATTTGCATTGGCTTTAGCCATTTGG	2536	Db	3782	ACATAGAAGAAAAGTCTCTGACTTGC	3841	Qy
2702	AGTGTGCAGTTAAACAGTTGTGATGCTACCAATATGGCGAGAAATTTTAGGACATACCA	2761	Qy	3617	ACATAGAAGAAAAGTCTCTGACTTGC	3676	Db
2537	AGTGTGCAGTTAAACAGTTGTGATGCTACCAATATGGCGAGAAATTTTAGGACATACCA	2596	Db	3842	CCAAAGGAAATGATTTTTTTCATCGAT	3901	Qy
2762	GGTTACACCGGATGACATCAATTTGAAGAGAGAAAGGAGAGAAAGTCAAAAAAAGAGA	2821	Qy	3677	CCAAAGGAAATGATTTTTTTCATCGAT	3736	Db
2597	GGTTACACCGGATGACATCAATTTGAAGAGAGAAAGGAGAGAAAGTCAAAAAAAGAGA	2656	Db	3902	CTTTGATCTGTAAAGACAAAACCGCTC	3961	Qy
2822	AGAAACAGGAAGAGAAACGATGCAGCAACGACATGGGTAAAAATACACATTTCCAG	2881	Qy	3737	CTTTGATCTGTAAAGACAAAACCGCTC	3796	Db
2657	AGAAACAGGAAGAGAAACGATGCAGCAACGACATGGGTAAAAATACACATTTCCAG	2716	Db	3962	CTTCCAAGAGTGTCTTACATTCATCAGT	4021	Qy
2882	TTAAGCATCAGTTTGGAAACAAAAGGTGAAGAGTACAGAGTACAGAGTATGTTGGTT	2941	Qy	3797	CTTCCAAGAGTGTCTTACATTCATCAGT	3856	Db
2717	TTAAGCATCAGTTTGGAAACAAAAGGTGAAGAGTACAGAGTACAGAGTATGTTGGTT	2776	Db	4022	CTCTGTCAAGAGCAATGGAGCTTTTGA	4081	Qy
2942	GGAGCTGGATTAGTAAACACTCATGTTTATAGGTTTGTCTCTAAATTCGCCAGGCAACTACTA	3001	Qy	3857	CTCTGTCAAGAGCAATGGAGCTTTTGA	3916	Db
2777	GGAGCTGGATTAGTAAACACTCATGTTTATAGGTTTGTCTCTAAATTCGCCAGGCAACTACTA	2836	Db	4082	CTTTGGAAATAGTTCTGTATACCGCTCT	4141	Qy
3002	ATGTGAATACAGAAAGTCTTTAGAGGACCAACCAAAATATATGGATGAAAAATPATGGATG	3061	Qy	3917	CTTTGGAAATAGTTCTGTATACCGCTCT	3976	Db
2837	ATGTGAATACAGAAAGTCTTTAGAGGACCAACCAAAATATATGGATGAAAAATPATGGATG	2896	Db	4142	TTCAGAAATAGCAATGAAAGCAATTTCT	4201	Qy
3062	AGTCAGATAAAGAAATGTTTCAGAGTCCCAAAAAAATAAATAATAGACCTGATTCTG	3121	Qy	3977	TTCAGAAATAGCAATGAAAGCAATTTCT	4036	Db
2897	AGTCAGATAAAGAAATGTTTCAGAGTCCCAAAAAAATAAATAATAGACCTGATTCTG	2956	Db	4202	TCPTGGAGCGTTTAAAGTGTGAGTTG	4261	Qy
3122	AAAAAGATAGGTTAAAGGTTTCAGATGCTGCAAAAGGAGCAGACCAAAATGAAATGGATA	3181	Qy	4037	TCPTGGAGCGTTTAAAGTGTGAGTTG	4096	Db
2957	AAAAAGATAGGTTAAAGGTTTCAGATGCTGCAAAAGGAGCAGACCAAAATGAAATGGATA	3016	Db				

Db	6257	TCCAAAGTCAACACTAGGAAAGCAATATTTCGAACACACCTGTGATGGTACAGCCAGGTG	6316	QY	7562	TCCAGTTCACAGGTTGTGGCTCAGATACAGGCTCAGCAAAAGTGGTGTGCCCCAGCAAAATCA	7621
QY	6482	CTCCTCAGCAAGTATGACTCAAAATCATCAGGGGGCAGCCCTGTCTCCACTGCAGTCCCG	6541	Db	7397	TCCAGTTCACAGGTTGTGGCTCAGATACAGGCTCAGCAAAAGTGGTGTGCCCCAGCAAAATCA	7456
Db	6317	CTCCTCAGCAAGTATGACTCAAAATCATCAGGGGGCAGCCCTGTCTCCACTGCAGTCCCG	6376	QY	7622	AACTCCAGTTACCTATCCAAAATTCAGCAAAAGCAGTCTGTGCAGACTCCACAGATTTCAGA	7681
QY	6542	CCCCTAAACAGGTTTCTCTCAACACCTGGGCAGAAAAGCTTAACTTCAGCAAACTGCCACTT	6601	Db	7457	AACTCCAGTTACCTATCCAAAATTCAGCAAAAGCAGTCTGTGCAGACTCCACAGATTTCAGA	7516
Db	6377	CCCCTAAACAGGTTTCTCTCAACACCTGGGCAGAAAAGCTTAACTTCAGCAAACTGCCACTT	6436	QY	7682	ATGTGGTTACAGTGCAGGCGAGCCAGTGTCAAGAGCAGTTGCAAAAGGGTTCAGCAACTCA	7741
QY	6602	CAAAATACAGTCTTTCAGGCTCAACAACCCCTCGCCCCCAACAGGACAAAGTGAAGTCA	6661	Db	7517	ATGTGGTTACAGTGCAGGCGAGCCAGTGTCAAGAGCAGTTGCAAAAGGGTTCAGCAACTCA	7576
Db	6437	CAAAATACAGTCTTTCAGGCTCAACAACCCCTCGCCCCCAACAGGACAAAGTGAAGTCA	6496	QY	7742	GGGATCAGCAGCAAAAAGAAAAGACAAACAGACATAGAAAATTAA-CGTGAACACACCCTCC	7800
QY	6662	CCATGGCTCAACTTACTCAGTTAACACAGGGCCACGGTGCATCAAGGTTTGACAGTAG	6721	Db	7577	GGGATCAGCAGCAAAAAGAAAAGACAAACAGACATAGAAAATTAAAGCGTGAACACACCCTCC	7636
Db	6497	CCATGGCTCAACTTACTCAGTTAACACAGGGCCACGGTGCATCAAGGTTTGACAGTAG	6556	QY	7801	AACTTCTTAATCAAAAGTTGAAATCAATCAGAAACAGGTGGTGTATGAAGCATATGCTGTA	7860
QY	6722	TAATTCAGGACAAGGTCAAACTACTGGACAGTTTGCAGTTGATACCTCAAGGGGTGACTG	6781	Db	7637	AACTTCTTAATCAAAAG-TGAAATCATTCAGAAACAGGTGGTGTATGAAGCATATGCTGTA	7695
Db	6557	TAATTCAGGACAAGGTCAAACTACTGGACAGTTTGCAGTTGATACCTCAAGGGGTGACTG	6616	QY	7861	ATAGAACATTTAAAAACAGAAAAAGAGCATGACTCCAGCTGAAAGAGAGAGATCAAGA	7920
QY	6782	TACTCCAGGCCAGGCCAGCAGCTAATGCAAGCTGCAATGCCAAATGGTACTGTTGAGC	6841	Db	7696	ATAGAACATTTAAAAACAGAAAAAGAGCATGACTCCAGCTGAAAGAGAGAGATCAAGA	7755
Db	6617	TACTCCAGGCCAGGCCAGCAGCTAATGCAAGCTGCAATGCCAAATGGTACTGTTGAGC	6676	QY	7921	ATGATTGTCTGTAAACAGAGTGTAAAGTATATTTTGGATAAGATAGATAAAGAAAGAAA	7980
QY	6842	GATTCTCTTTTACCCCATTTGCAACACAGCCACACAGCCAGCCAGCCACACACACCACTG	6901	Db	7756	ATGATTGTCTGTAAACAGAGTGTAAAGTATATTTTGGATAAGATAGATAAAGAAAGAAA	7815
Db	6677	GATTCTCTTTTACCCCATTTGCAACACAGCCACACAGCCAGCCAGCCACACACCACTG	6736	QY	7981	CAGGCAGCAAAAACCGRAAGCGTGAAGAGAGTGTGGAGCAGAAACCTAGCAAGCAGAAT	8040
QY	6902	TTTCCACGACAGCAGCGGTACAGGTGAACAAAAGGCAGAGTAAACTGTCAACCCAGATGC	6961	Db	7816	CAGGCAGCAAAAACCGRAAGCGTGAAGAGAGTGTGGAGCAGAAACCTAGCAAGCAGAAT	7875
Db	6737	TTTCCACGACAGCAGCGGTACAGGTGAACAAAAGGCAGAGTAAACTGTCAACCCAGATGC	6796	QY	8041	GCCACTAAGCTGTGAGTCTGCTCTTCAAGCAAAAAGAGCAGCTCAGAGCCGAGATCCTG	8100
QY	6962	AGGTACATACAGAAAACCTCCACACAGCTCAGTCAATCAAGTGTGGTCCACCAAG	7021	Db	7876	GCCACTAAGCTGTGAGTCTGCTCTTCAAGCAAAAAGAGCAGCTCAGAGCCGAGATCCTG	7935
Db	6797	AGGTACATACAGAAAACCTCCACACAGCTCAGTCAATCAAGTGTGGTCCACCAAG	6856	QY	8101	AAGAGAGAGCCTCCTGGACAGGATCTGCAAAATTTGAAGTGCAGGAGAGCTGAAGAGA	8160
QY	7022	CCAGCCACAGACTGCTCAGCCTTTCAGCTCGGCCCCAGCCCAACCCAGCCAGCCAGTCCC	7081	Db	7936	AAGAGAGAGCCTCCTGGACAGGATCTGCAAAATTTGAAGTGCAGGAGAGCTGAAGAGA	7995
Db	6857	CCAGCCACAGACTGCTCAGCCTTTCAGCTCGGCCCCAGCCCAACCCAGCCAGTCCC	6916	QY	8161	GACCTGAAAATTAAGAAAAGAAAAGACCTGATGCAGTTGGCTCAGGCCACAGCAGTAGCT	8220
QY	7082	CAGCTCAGCCTGAAGTTTCAGACTCAGCCTGAAGTTTCAGACCCCAACAACTGTTTCATCCC	7141	Db	7996	GACCTGAAAATTAAGAAAAGAAAAGACCTGATGCAGTTGGCTCAGGCCACAGCAGTAGCT	8055
Db	6917	CAGCTCAGCCTGAAGTTTCAGACTCAGCCTGAAGTTTCAGACCCCAACAACTGTTTCATCCC	6976	QY	8221	GCACCTTGCCTCCAGTGACACACAGTTCCTCCAGCCCTCCAGCCCTCCACCTTCACCT	8280
QY	7142	ATGTCCCTTCTGAAGCAACACCACCGACAGTCAATCAAGCCCAAGTTCGAGCAG	7201	Db	8056	GCACCTTGCCTCCAGTGACACACAGTTCCTCCAGCCCTCCAGCCCTCCACCTTCACCT	8115
Db	6977	ATGTCCCTTCTGAAGCAACACCACCGACAGTCAATCAAGCCCAAGTTCGAGCAG	7036	QY	8281	CCCCCTCCAGTGTGTGCAACACACAGGCTTCTGTCCAGCCCTCCAGCCCTCCACCTTCACCT	8340
QY	7202	AGTCTCAGCCTCAAAAGTAAATGTCCAAGGACAGTCTCCTGTTGCTGTCCAAAGTCCATCAC	7261	Db	8116	CCCCCTCCAGTGTGTGCAACACACAGGCTTCTGTCCAGCCCTCCAGCCCTCCACCTTCACCT	8175
Db	7037	AGTCTCAGCCTCAAAAGTAAATGTCCAAGGACAGTCTCCTGTTGCTGTCCAAAGTCCATCAC	7096	QY	8341	TCCCAGAGAGGAAAGCGGAGAGGAAAAGACTCCAGCTCAAAAGTCCAAAGAAAAGAAA	8400
QY	7262	AGACTCGAATAGTCCATCACTCCCACTGTCTCTGAGCAACAACTCCAGGTTTC	7321	Db	8176	TCCCAGAGAGGAAAGCGGAGAGGAAAAGACTCCAGCTCAAAAGTCCAAAGAAAAGAAA	8235
Db	7097	AGACTCGAATAGTCCATCACTCCCACTGTCTCTGAGCAACAACTCCAGGTTTC	7156	QY	8401	ATGATCTCTACTACTCTCAAGGAACTTAAGAAAGACAAAGCTTTACTGTATCTGTA	8460
QY	7322	AGACTACAACTCAACACCGATTCCAATTCACACACATACATCTCTTCAGATACCTTCCC	7381	Db	8236	ATGATCTCTACTACTCTCAAGGAACTTAAGAAAGACAAAGCTTTACTGTATCTGTA	8295
Db	7157	AGACTACAACTCAACACCGATTCCAATTCACACACATACATCTCTTCAGATACCTTCCC	7216	QY	8461	ACGCTTATGATGATCTTAAATTTTATATTTGGCTGTGATCGGTGTGAGATTTGTTACCAT	8520
QY	7382	AAGGCCAGCCAGTCAACACCCAGGTACAGTCTTCAACTCAAACTCTTTCATCAGGAC	7441	Db	8296	ACGCTTATGATGATCTTAAATTTTATATTTGGCTGTGATCGGTGTGAGATTTGTTACCAT	8355
Db	7217	AAGGCCAGCCAGTCAACACCCAGGTACAGTCTTCAACTCAAACTCTTTCATCAGGAC	7276	QY	8521	GGGCGCTCGTGGCATCTTGCAAAAGTGAGGCGAGCTCATTTGATGAGTATGCTGTCCA	8580
QY	7442	AACTTTAAATCAAGTTAGTGTTCATCCCATCCGTCCTCAGCTACAAATACAGCAGC	7501	Db	8356	GGGCGCTCGTGGCATCTTGCAAAAGTGAGGCGAGCTCATTTGATGAGTATGCTGTCCA	8415
Db	7277	AACTTTAAATCAAGTTAGTGTTCATCCCATCCGTCCTCAGCTACAAATACAGCAGC	7336	QY	8581	CAGTCCAGTCAACAGAGGATGCCATGACAGTGTCTACGCCACTTAACAGAGAAGGATAT	8640
QY	7502	CACAGCCCAAGTCAATTTGCTGCTCAGCTGCAACAAAGTCCAGGTTCTCTCTCAGA	7561	Db	8416	CAGTCCAGTCAACAGAGGATGCCATGACAGTGTCTACGCCACTTAACAGAGAAGGATAT	8475
Db	7337	CACAGCCCAAGTCAATTTGCTGCTCAGCTGCAACAAAGTCCAGGTTCTCTCTCAGA	7396				

QY 8641 GAGGGTTGAAGGGTGTCTCGTTCCCTTACAGGCCCATAAAGATGCCCTGGCCCTTCCTTT 8700
Db 8476 GAGGGTTGAAGGGTGTCTCGTTCCCTTACAGGCCCATAAAGATGCCCTGGCCCTTCCTTT 8535
QY 8701 GAACAGTAGACCCCTAATGATGCACAGATTTATATGTTGTTTAAAGAACTATGGAC 8760
Db 8536 GAACAGTAGACCCCTAATGATGCACAGATTTATATGTTGTTTAAAGAACTATGGAC 8595
QY 8761 CTTGGCCACCATGGAAGAAAGATACAAAGACGATATTTATGAAAGCTGACGGAATTTGTG 8820
Db 8596 CTTGGCCACCATGGAAGAAAGATACAAAGACGATATTTATGAAAGCTGACGGAATTTGTG 8655
QY 8821 GCAGATATGACCAAAATTTTGTATTAACCTCTCTACTACAAATCCAAAGTCCCAATTT 8880
Db 8656 GCAGATATGACCAAAATTTTGTATTAACCTCTCTACTACAAATCCAAAGTCCCAATTT 8715
QY 8881 TACCAGTGTGACAGAGTTCTCGAATCATTTCTTGTGACAGAAATGAAAGGCTTCAAAGCT 8940
Db 8716 TACCAGTGTGACAGAGTTCTCGAATCATTTCTTGTGACAGAAATGAAAGGCTTCAAAGCT 8775
QY 8941 AGCAGTCTCATACAAACAAATGCAGTCTCAGCTTCTTAAAGTTCAGCGTGTAAACCT 9000
Db 8776 AGCAGTCTCATACAAACAAATGCAGTCTCAGCTTCTTAAAGTTCAGCGTGTAAACCT 8835
QY 9001 ACATATAACACAGCAAGATCTGTTCTCTCAACTATTTTAAATTAAGGACCCAGATG 9060
Db 8836 ACATATAACACAGCAAGATCTGTTCTCTCAACTATTTTAAATTAAGGACCCAGATG 8895
QY 9061 TTTTGTAGTCAAGGCTTCTGACAGACTTGACCTTAACTTCGTTTATTGTTCAATAACAG 9120
Db 8896 TTTTGTAGTCAAGGCTTCTGACAGACTTGACCTTAACTTCGTTTATTGTTCAATAACAG 8955
QY 9121 TCCAAATATATTTCTGGCCAAATTTTGTCCACGACGACGACGACGACGACGACGAC 9180
Db 8956 TCCAAATATATTTCTGGCCAAATTTTGTCCACGACGACGACGACGACGACGACGAC 9015
QY 9181 CCATATATCTGTGACAGATCAGATGTTTACTTATTGTCGACAGGACGACGACGACGAC 9240
Db 9016 CCATATATCTGTGACAGATCAGATGTTTACTTATTGTCGACAGGACGACGACGACGAC 9075
QY 9241 TATTGAAAAAAGAAAAAGAAAGCAAGAAAAAGATACATATGGGTCAAGTGTAACT 9300
Db 9076 TATTGAAAAAAGAAAAAGAAAGCAAGAAAAAGATACATATGGGTCAAGTGTAACT 9135
QY 9301 CCATGGAATGCAGCTCTGCTTCTCAGTGAAGAGCTGGTTAGAGTCTCACGAAAAAC 9360
Db 9136 CCATGGAATGCAGCTCTGCTTCTCAGTGAAGAGCTGGTTAGAGTCTCACGAAAAAC 9195
QY 9361 TTTTGTAGTCTATTTATTTATTTGTCGACGACGCTTTTATTGCTGCGCTCATTTG 9420
Db 9196 TTTTGTAGTCTATTTATTTATTTGTCGACGACGCTTTTATTGCTGCGCTCATTTG 9255
QY 9421 TCAGCTAAGTATTTTCTTATAAATCCAGCCCGGTTACATATAATCATCTGTATCTT 9480
Db 9256 TCAGCTAAGTATTTTCTTATAAATCCAGCCCGGTTACATATAATCATCTGTATCTT 9315
QY 9481 ATCATGATTCCTGTAGTAAAGTACAGACGACCTCTAGATGCTTTTCTTCTATGAA 9540
Db 9316 ATCATGATTCCTGTAGTAAAGTACAGACGACCTCTAGATGCTTTTCTTCTATGAA 9375
QY 9541 AGGAGCTGCTATGACATGTCACACACACACACACACACACACACACACACACACACAC 9600
Db 9376 AGGAGCTGCTATGACATGTCACACACACACACACACACACACACACACACACACACAC 9435
QY 9601 GTTTCAGTGTAGATTTAAATTAAGCTTGCATTAAGGTTGGGCTAAGTGGTCCCTGGCTAC 9660
Db 9436 GTTTCAGTGTAGATTTAAATTAAGCTTGCATTAAGGTTGGGCTAAGTGGTCCCTGGCTAC 9495
QY 9661 AGACTCTGTGCTTGAATATAACAGTACAAATTTCTCAATTAATCTGTCACACAGGCTAAG 9720
Db 9496 AGACTCTGTGCTTGAATATAACAGTACAAATTTCTCAATTAATCTGTCACACAGGCTAAG 9555
QY 9721 TGAGTAAATCTATTGTAAGGTATCTGTTGTTGTAACATTTGTCAGATTTCTAATTTTTTT 9780

Db 9556 TCAGTAAATCTATTGTAAGGTATCTGTTGTTAAACATTTGTCAGATTTCTAATTTTTTT 9615
QY 9781 CTTTGTATTAAAAATCAACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9840
Db 9616 CTTTGTATTAAAAATCAACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9675
QY 9841 CTCACACAAAAAATAAAAAA 9865
Db 9676 CTCACACAAAAAATAAAAAA 9700

RESULT 3
AAL51568
ID AAL51568 standard; DNA; 8280 BP.
XX
AC AAL51568;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID NO 51.
XX
KW Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN WO2003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US21179.
XX
PR 22-JUN-2001; 2001US-300518P.
PR 29-JUN-2001; 2001US-301787P.
PR 29-JUN-2001; 2001US-301792P.
PR 29-JUN-2001; 2001US-301892P.
PR 29-JUN-2001; 2001US-301893P.
PR 06-JUL-2001; 2001US-303405P.
PR 06-JUL-2001; 2001US-303442P.
PR 15-MAR-2002; 2002US-364438P.
XX
PA (INCV-) INCVTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;
PI Borowski ML, Yao MG, Walla NK, Bandman O, Lal PG, Becha SD;
PI Lee SY, Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjian Y;
PI Lu Y;
XX
DR WPI; 2003-201420/19.
DR P-PSDB; AAO16418.
XX
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS)
XX
PS Claim 12; Page 289-292; 312pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders

CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC the invention are useful for creating transgenic animals to model human
CC disease. The present DNA sequence encodes a human nucleic acid-associated
CC protein of the invention.

Sequence 8280 BP; 2841 A; 1761 C; 1873 G; 1805 T; 0 other;

Query Match	73.5%;	Score 7255.2;	DB 25;	Length 8280;
Best Local Similarity	90.9%;	Pred. No. 0;		
Matches 8088; Conservative	0;	Mismatches 18;	Indels 792;	Gaps 3;

259	Qy	ATGGTCTCCGAGAGGAGGAGGAGGAGCGCGCGAGGAGACCCAGGATTCGTGAG	318
	Db	1 ATGGTCTCCGAGAGGAGGAGGAGGAGCGCGCGAGGAGACCCAGGATTCGTGAG	60
319	Qy	GAGCAGCAGGAGGATGAGATGGAAGAGGACGACGATGACTCCGATTATCCGAGGAGATG	378
61	Db	GACGACGAGGAGATGAGATGGAAGAGGACGACGATGACTCCGATTATCCGAGGAGATG	120
379	Qy	GAAGCAGCAGCAGCAGCGCCAGCTTACTGCACGGAAAGCAGCTTCAGGAGCCATAGTACC	438
121	Db	GAAGCAGCAGCAGCAGCGCCAGCTTACTGCACGGAAAGCAGCTTCAGGAGCCATAGTACC	180
439	Qy	TACAGCAGCACTCCAGGTAGCGCAAAACCAAGAGTACATCGGCTCGTTCTCCTATATTG	498
181	Db	TACAGCAGCACTCCAGGTAGCGCAAAACCAAGAGTACATCGGCTCGTTCTCCTATATTG	240
499	Qy	GAAGAAAAGACATCCCGCCCTTGAATTTCCCAAGTCCTCTGAGGATTAATGTGCCT	558
241	Db	GAAGAAAAGACATCCCGCCCTTGAATTTCCCAAGTCCTCTGAGGATTAATGTGCCT	300
559	Qy	AATGAGCATATATGAATGTCAATGGCATTTACGAGGTACTGCGGAACCTTGGCACTGTT	618
301	Db	AATGAGCATATATGAATGTCAATGGCATTTACGAGGTACTGCGGAACCTTGGCACTGTT	360
619	Qy	TTGAGATTACTCCTTTTCGCTTTGAGGACTTTTGTGCGACTCTCTGTGAGCCCAAGAGCAG	678
361	Db	TTGAGATTACTCCTTTTCGCTTTGAGGACTTTTGTGCGACTCTCTGTGAGCCCAAGAGCAG	420
679	Qy	TGCACACTCATGGCAGAGATGCATTTGTGCTTTTGAAGACAGTTCTGGGTGAAGAAGAC	738
421	Db	TGCACACTCATGGCAGAGATGCATTTGTGCTTTTGAAGACAGTTCTGGGTGAAGAAGAC	480
739	Qy	ACTTCCAAATACTACTTTGGACCTGCTGATCTGAAAGATAGCCTTAATCCACACACTGAT	798
481	Db	ACTTCCAAATACTACTTTGGACCTGCTGATCTGAAAGATAGCCTTAATCCACACACTGAT	540
799	Qy	TTTCATAGATGGATGACGTGGCCAGAGAGTGCTGCGGGTGTACTGTGAGAGTGATAAGGAG	858
541	Db	TTTCATAGATGGATGACGTGGCCAGAGAGTGCTGCGGGTGTACTGTGAGAGTGATAAGGAG	600
859	Qy	TACCATACGTTCTTCCTTTACCAAGAGCAGAGGACTACCCATATGACACAGTAGAGAAC	918
601	Db	TACCATACGTTCTTCCTTTACCAAGAGCAGAGGACTACCCATATGACACAGTAGAGAAC	660
919	Qy	AGATCAAAAGTTCTACAGTTTCTAGTCGATCAGTTTCTTACACAAATATTGCTCCGAGAG	978
661	Db	AGATCAAAAGTTCTACAGTTTCTAGTCGATCAGTTTCTTACACAAATATTGCTCCGAGAG	720
979	Qy	GAATTTGATGTCGTAAGGGGTGATACAGTATGATGACCATGTAGGGTTTGTCCAAACTT	1038
721	Db	GAATTTGATGTCGTAAGGGGTGATACAGTATGATGACCATGTAGGGTTTGTCCAAACTT	780
1039	Qy	GGGGATTTGCTTTTGGTGTGAGACATGTTTCAGCAGTATACCATTTGGAATGTGTGAAGCCA	1098
781	Db	GGGGATTTGCTTTTGGTGTGAGACATGTTTCAGCAGTATACCATTTGGAATGTGTGAAGCCA	840
1099	Qy	CCTCTTGAGGAGGTCCAGAGGACGAGTGGCAGTGTGAAGTCTCTGTAGCACACAAGGTG	1158
841	Db	CCTCTTGAGGAGGTCCAGAGGACGAGTGGCAGTGTGAAGTCTCTGTAGCACACAAGGTG	900
1159	Qy	CCTGGTGTGACTGACTGTTGTGCTGAAATCCAAAAATAAACCATATATTCGACATGAA	1218

901	DB	CTGGTGTGACTGACTGTGTTCTGCTGAATCCAAAAAATAAACCATATATATTTCGACATGAA	960
1219	QY	CCATATTGGATATGATAGAAAGTCGGAGGAAATACTGGTTCTTGAAACCGAGACTCATATA	1278
961	DB	CCATATTGGATATGATAGAAAGTCGGAGGAAATACTGGTTCTTGAAACCGAGACTCATATA	1020
1279	QY	GAAGAGATACAGAAATGAAATGAAAGGAAATTTGGTATTACAGACAAAGGTCCAA	1338
1021	DB	GAAGAGATACAGAAATGAAATGAAAGGAAATTTGGTATTACAGACAAAGGTCCAA	1080
1339	QY	CTTGCAAGATTAATTGACTGTCTAGACAAAGATTATTGGGAAGCAGAACTCTGC AAAATT	1398
1081	DB	CTTGCAAGATTAATTGACTGTCTAGACAAAGATTATTGGGAAGCAGAACTCTGC AAAATT	1140
1399	QY	CTAGAGAAATCGGTGAAGAAATCCACCCACACATGGACATAACTGAAGACCTGACCAAT	1458
1141	DB	CTAGAGAAATCGGTGAAGAAATCCACCCACACATGGACATAACTGAAGACCTGACCAAT	1200
1459	QY	AAGGCTCGGGCAGTAACAAATCCCTTCTGCGGCGCAGCTAAATGAAGAAATTTTGAATCC	1511
1201	DB	AAGGCTCGGGCAGTAACAAATCCCTTCTGCGGCGCAGCTAAATGAAGAAATTTTGAATCC	1260
1519	QY	ATAAGACCCAAAGGGAGACATTTGATATGTTTAAAGCCCGAGAGAAACACAGAAAGAC	1578
1261	DB	ATAAGACCCAAAGGGAGACATTTGATATGTTTAAAGCCCGAGAGAAACACAGAAAGAC	1320
1579	QY	AGAATTCAGACTCAGAAATGACTCTAAAGATGCTGAGAAAAACAGAGAGAATTTGAAGAC	1638
1321	DB	AGAATTCAGACTCAGAAATGACTCTAAAGATGCTGAGAAAAACAGAGAGAATTTGAAGAC	1380
1639	QY	CAGTCCCTTTGAAAAAGACAGTGCACGACAAACACACAGATGATGACCTGAGCAGGAAAA	1698
1381	DB	CAGTCCCTTTGAAAAAGACAGTGCACGACAAACACACAGATGATGACCTGAGCAGGAAAA	1440
1699	QY	TCCTGAGGAGCCACAGAGTTGGGGATAAGGTAACCTGTGTGCACAAATCTTGCGGAC	1758
1441	DB	TCCTGAGGAGCCACAGAGTTGGGGATAAGGTAACCTGTGTGCACAAATCTTGCGGAC	1500
1759	QY	AACACAACAATGCACTTCAGAGAGACTAGTCCCTCTGAAGGAGGAGCCCTGTGGGG	1818
1501	DB	AACACAACAATGCACTTCAGAGAGACTAGTCCCTCTGAAGGAGGAGCCCTGTGGGG	1560
1819	QY	TGTCCTCAGAAACCCCGATAGCAGCAACATGGCAGAGAGAGGTGGCATCTGAGCTC	1878
1561	DB	TGTCCTCAGAAACCCCGATAGCAGCAACATGGCAGAGAGAGGTGGCATCTGAGCTC	1620
1879	QY	CCCAGGATGTGCCAGAGAAACCTTAACAGACATGTGAGAGCAGTAACACTAGTGTCTACC	1938
1621	DB	CCCAGGATGTGCCA	1635
1939	QY	ACTACCTCCATCCAGCCATAATCTGGAAAAACAGTAACAGCAGCAGTGAATTAATTTCTTC	1998
1636	DB	-----	1635
1999	QY	CAGAGTGAATCTGCTAAGGCAGCTGATGCTCTGAAATGGAGAAAGAGAATCTCATACA	2058
1636	DB	-----	1635
2059	QY	CCGTGCTCTATTTCAGGAAGAGATAGTAGGTGATTTACATCGGAGAGAGTCCACCGGGGAG	2118
1636	DB	-----GTAGGTGATTTCAAATCGGAGAGAGTCCACCGGGGAG	1671
2119	QY	CTAAGTGAATCTCTGAGCTGGAAAAGGAGCATCTGCTCAACTCGAATCATCACCCAGA	2178
1672	DB	CTAAGTGAATCTCTGAGCTGGAAAAGGAGCATCTGCTCAACTCGAATCATCACCCAGA	1731
2179	QY	TTGCGGAATCCAGATAGCAAACTTTAGTCAGCTGAAGGCCACGACAGGTGGCAGCGCGTGCA	2238
1732	DB	TTGCGGAATCCAGATAGCAAACTTTAGTCAGCTGAAGGCCACGACAGGTGGCAGCGCGTGCA	1791
2239	QY	CATGAGCAATAAATTTAATTAAGGAGGGCAAGAGGTACTGGTGTAGTTAACTCTCAAGGA	2298
1792	DB	CATGAGCAATAAATTTAATTAAGGAGGGCAAGAGGTACTGGTGTAGTTAACTCTCAAGGA	1851

QY	2299	GAAATTTACGGTTGAGCACCAAAAAAGGAGGTGATCATGAAAGGAAATATCAACAATTTAT	2358	3379	AAAACAAAATCATCCAACTAGATGACATCTTTGAAAGGAGAATTAACACAGTTTACACTG	3438
Db	1852	GAAATTTACGGTTGAGCACCAAAAAAGGAGGTGATCATGAAAGGAAATATCAACAATTTAT	1911	2932	AAAACAAAATCATCCAACTAGATGACATCTTTGAAAGGAGAATTAACACAGTTTACACTG	2991
QY	2359	TTTAAATTTGGTCAACAAGGAGATCGGCTACCAACAAATACATCCACCAATTTCA	2418	3439	GAAGAAAAACAGGACTCGAAAAAATCAAGTTGAGGGTGAATTAAGGGTATAGGAAG	3498
Db	1912	TTTAAATTTGGTCAACAAGGAGATCGGCTACCAACAAATACATCCACCAATTTCA	1971	2992	GAAGAAAAACAGGACTCGAAAAAATCAAGTTGAGGGTGAATTAAGGGTATAGGAAG	3051
QY	2419	TTTGGCTTTGAATAAGCACCAGCACAGAGAGACCATGATAAGAGAAGGCATCTTGCACAT	2478	3499	ACTTCTACAAATTTCTTCAAAAAATCTCTGTAATCACCAAGTAAATACGAAAGCAAAAGAA	3558
Db	1972	TTTGGCTTTGAATAAGCACCAGCACAGAGAGACCATGATAAGAGAAGGCATCTTGCACAT	2031	3052	ACTTCTACAAATTTCTTCAAAAAATCTCTGTAATCACCAAGTAAATACGAAAGCAAAAGAA	3111
QY	2479	AAGTTCTGTGACTCCAGCAGAGAGTTCAAAATGGAACGGTTCTGTCCATGGGTCCAAA	2538	3559	GGGTGTCAGAGTGCATGAGACAGAAACAGACAGCCCAAAATCAATATATGATCAACCT	3618
Db	2032	AAGTTCTGTGACTCCAGCAGAGAGTTCAAAATGGAACGGTTCTGTCCATGGGTCCAAA	2091	3112	GGGTGTCAGAGTGCATGAGACAGAAACAGACAGCCCAAAATCAATATATGATCAACCT	3171
QY	2539	GTTCCTTACCATATCTACTCTGAGACTGACTATCACCCCAATTAGAAAACAACATCCCTTCA	2598	3619	GAGGACTTGATTTCCAGGATGTTTCACAAAGTGATTCCTCAGTTCTTGAATGAGTGATCCT	3678
Db	2092	GTTCCTTACCATATCTACTCTGAGACTGACTATCACCCCAATTAGAAAACAACATCCCTTCA	2151	3172	GAGGACTTGATTTCCAGGATGTTTCAGAAAGTGATTCCTCAGTTCTTGAATGAGTGATCCT	3231
QY	2599	TCCTTTTTCATCCCACTGGGCATCACATAGGCAAAATTTGGATCAAGCAGTTTCAGATG	2658	3679	AGTCATACCAACCAAACTTTATCCAAAAGATCGAGTGTAGATGATGTCTCCATTCGG	3738
Db	2152	TCCTTTTTCATCCCACTGGGCATCACATAGGCAAAATTTGGATCAAGCAGTTTCAGATG	2211	3232	AGTCATACCAACCAAACTTTATCCAAAAGATCGAGTGTAGATGATGTCTCCATTCGG	3291
QY	2659	TGTAGCAAAACCCAGAGAAATTTGCATTTGGCTTTAGCCATTTTGGAGTGTGCAGTTAAACCA	2718	3739	AGCCAGAAAACAAAATGTCGGAACAAAATTTCCATTTGAAATGACATAGAGAAGAAAGTC	3798
Db	2212	TGTAGCAAAACCCAGAGAAATTTGCATTTGGCTTTAGCCATTTTGGAGTGTGCAGTTAAACCA	2271	3292	AGCCAGAAAACAAAATGTCGGAACAAAATTTCCATTTGAAATGACATAGAGAAGAAAGTC	3351
QY	2719	GTTCGTATGCTACCAATATGGCGAGAAATTTTAGGACATACCAAGTTACACCGGATGACA	2778	3799	TCGTGACTTGCCTAGTAGAGCCAGAACCCACTTAAGAGTAAACACCAAGGAAATGATTTT	3858
Db	2272	GTTCGTATGCTACCAATATGGCGAGAAATTTTAGGACATACCAAGTTACACCGGATGACA	2331	3352	TCGTGACTTGCCTAGTAGAGCCAGAACCCACTTAAGAGTAAACACCAAGGAAATGATTTT	3411
QY	2779	TCAATTTGAAGAGAGAGAGAGAAAGTCAAAAAAAGAGAGAAACAGGAAGAGAA	2838	3859	TTCATCGATGACTCTAAACTAGCCAGTGCAGATGATATTGGTACTTTGATCTCTAAGAAC	3918
Db	2332	TCAATTTGAAGAGAGAGAGAGAAAGTCAAAAAAAGAGAGAAACAGGAAGAGAA	2391	3412	TTCATCGATGACTCTAAACTAGCCAGTGCAGATGATATTGGTACTTTGATCTCTAAGAAC	3471
QY	2839	GAACGATGCAGAACGACATGGGTAAATACACATTTCCAGTTAAGCATCAGGTTTGG	2898	3919	AAAAACCGCTCATACAGGAGGAAAGTGACACCATTTGTTCTTCTCCAAAGAGTGCTTTA	3978
Db	2392	GAACGATGCAGAACGACATGGGTAAATACACATTTCCAGTTAAGCATCAGGTTTGG	2451	3472	AAAAACCGCTCATACAGGAGGAAAGTGACACCATTTGTTCTTCTCCAAAGAGTGCTTTA	3531
QY	2899	AAACAAAAGGTGAACAGTACAGAGTACAGAGATGGTGGTGGAGCTGGATTTAGTAA	2958	3979	CATTATCAGTGCCTAAAAGTACCAATGACAGAGATGCCACACTCTGTCAAGAGCAATG	4038
Db	2452	AAACAAAAGGTGAACAGTACAGAGTACAGAGATGGTGGTGGAGCTGGATTTAGTAA	2511	3532	CATTATCAGTGCCTAAAAGTACCAATGACAGAGATGCCACACTCTGTCAAGAGCAATG	3591
QY	2959	ACTCATGTTTATAGGTTTCTCCTAAATTTCCAGGCAATACTAATGTGAATTTACAGAAAG	3018	4039	GACTTTGAAAGGAAAACCTGGGATGTGACTCTGAATCTAATAGCACTTTTGGAAAAATAGTTCT	4098
Db	2512	ACTCATGTTTATAGGTTTCTCCTAAATTTCCAGGCAATACTAATGTGAATTTACAGAAAG	2571	3592	GACTTTGAAAGGAAAACCTGGGATGTGACTCTGAATCTAATAGCACTTTTGGAAAAATAGTTCT	3651
QY	3019	TCGTTAGAGGAACCAAAAAATATATGGATGAAATATGGATGAGTGCAGATAAAGAAAA	3078	4099	GATACCGTGTCTATTCCAGGATAGCAGTGAAGAGATATGATTTGTTTCCAGATAGCAATGAA	4158
Db	2572	TCGTTAGAGGAACCAAAAAATATATGGATGAAATATGGATGAGTGCAGATAAAGAAAA	2631	3652	GATACCGTGTCTATTCCAGGATAGCAGTGAAGAGATATGATTTGTTTCCAGATAGCAATGAA	3711
QY	3079	TGTTACAGAGTCCAAAAAATAAATAATAGAGCCTGATCTCAAAAAAGATGAGGTAA	3138	4159	AGCATTTCTGAACAGTTTCAGAACTCGAGAACAGAGATGTTGAAGTCTTGGAGCCGTTAAAG	4218
Db	2632	TGTTACAGAGTCCAAAAAATAAATAATAGAGCCTGATCTCAAAAAAGATGAGGTAA	2691	3712	AGCATTTCTGAACAGTTTCAGAACTCGAGAACAGAGATGTTGAAGTCTTGGAGCCGTTAAAG	3771
QY	3139	GGTTTCAGATGCTGCAAAAGGAGCAGACCAAAAATGAAATGGATGATCTCAAAAGATTTACTGAG	3198	4219	TGTGAGTTGGTTTCTGGTGGTCCACTGGAAACTGTGAGACAGGCTGCCGGTCAAGGGG	4278
Db	2692	GGTTTCAGATGCTGCAAAAGGAGCAGACCAAAAATGAAATGGATGATCTCAAAAGATTTACTGAG	2751	3772	TGTGAGTTGGTTTCTGGTGGTCCACTGGAAACTGTGAGACAGGCTGCCGGTCAAGGGG	3831
QY	3199	AGAAGGCCAAGATGTGAAGGAGCTCTTAGATCTGACAGTGAATAAACCCCTGCAAGGAA	3258	4279	ACTGAAGCAAAATGGTAAAAAACCAACTCAGCAGAGAAATTTAGAGGAGAGACCAAGTTAAT	4338
Db	2752	AGAAGGCCAAGATGTGAAGGAGCTCTTAGATCTGACAGTGAATAAACCCCTGCAAGGAA	2811	3832	ACTGAAGCAAAATGGTAAAAAACCAACTCAGCAGAGAAATTTAGAGGAGAGACCAAGTTAAT	3891
QY	3259	GAACCAATGGAATGAGACCATGACATGAAAAACAGAGTACATGTAATTTGTCAGGAGAT	3318	4339	AAATCTAGTGAATCAATTAAGCTTAAAAATACCCTGACAAAAGAGATTAATGAAATCGA	4398
Db	2812	GAACCAATGGAATGAGACCATGACATGAAAAACAGAGTACATGTAATTTGTCAGGAGAT	2871	3892	AAATCTAGTGAATCAATTAAGCTTAAAAATACCCTGACAAAAGAGATTAATGAAATCGA	3951
QY	3319	TCTCAAGTATGATGTGGTCAATGTTAGTGAGGGTTTTTCATCTAAGGACTAGTTTACAAAAAG	3378	4399	GAGTCTGAAAAAGAAAGGACAGACAAAGTACATTTCAATTAATGAAAAAGATAATAA	4458
Db	2872	TCTCAAGTATGATGTGGTCAATGTTAGTGAGGGTTTTTCATCTAAGGACTAGTTTACAAAAAG	2931	3952	GAGTCTGAAAAAGAAAGGACAGACAAAGTACATTTCAATTAATGAAAAAGATAATAA	4011
				4459	CCCAAAATATATTTGAAAGGTGAATGCTTTGAAAGAAATTTCTGAGAGTAGAGTAGAAGT	4518

|||||
4012 CCCAAATATATTTGAAAGGTGAATGCTTGAAGAAATTTCTGAGAGTAGTAGTAAAGT 4071 Db
QY
4519 GGTAAATGTTGAACCAAGGTTAAATATATAAATAAATAATCCCTGAGAATGATATATAA 4578 QY
Db
4072 GGTAAATGTTGAACCAAGGTTAAATATATAAATAAATAATCCCTGAGAATGATATATAA 4131 Db
QY
4579 TCATTTGACTGTTAAAGAAATCTGCTATTAAGGCCATTCAATTAATGGTGATGTCATCATGGAA 4638 QY
Db
4132 TCATTTGACTGTTAAAGAAATCTGCTATTAAGGCCATTCAATTAATGGTGATGTCATCATGGAA 4191 Db
QY
4639 GATTTTAAATGAAGAAACAGCTCCGAAACAAAATCGCATTTGCTGATGTTCTTTCAGATGCT 4698 QY
Db
4192 GATTTTAAATGAAGAAACAGCTCCGAAACAAAATCGCATTTGCTGATGTTCTTTCAGATGCT 4251 Db
QY
4699 GAAGGTAACTACCGAGATAGCTTTGAGACCCCTGCCATCAACCAAGAGTCTGACAGTACA 4758 QY
Db
4252 GAAGGTAACTACCGAGATAGCTTTGAGACCCCTGCCATCAACCAAGAGTCTGACAGTACA 4311 Db
QY
4759 CAGAGGACCAACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAGAT 4818 QY
Db
4312 CAGAGGACCAACCCCTCAGCATCTTGTCCAGAAAGCAATTCAGTTAATCAGGTAGAAGAT 4371 Db
QY
4819 ATGGAATATGAACCTCAGAAAGTTAAGAAAGTTACTTCATCACCTATTACTTCTGGAAGAG 4878 QY
Db
4372 ATGGAATATGAACCTCAGAAAGTTAAGAAAGTTACTTCATCACCTATTACTTCTGGAAGAG 4431 Db
QY
4879 GAATCTAATCTCAGTAATGACTTTATGATGAATGGTCTGCCCATCAACAAAATGAA 4938 QY
Db
4432 GAATCTAATCTCAGTAATGACTTTATGATGAATGGTCTGCCCATCAACAAAATGAA 4491 Db
QY
4939 AATGTCAATGGAGAAATCTAAAAGAAAAACCGTCTATCAGAAAGTCAACACGATGACCTCC 4998 QY
Db
4492 AATGTCAATGGAGAAATCTAAAAGAAAAACCGTCTATCAGAAAGTCAACACGATGACCTCC 4551 Db
QY
4999 ACAGTGGCCACAGAAATCAAAATCTGTGATCAAGGTAGAAAAAGGGGATAGCAAACTGTG 5058 QY
Db
4552 ACAGTGGCCACAGAAATCAAAATCTGTGATCAAGGTAGAAAAAGGGGATAGCAAACTGTG 4611 Db
QY
5059 GTTTCCTCCACAGAAAAATGTCAAAAATCCACTGTCAACAAACCCACTACAAACAGTGAAC 5118 Db
Db
4612 GTTTCCTCCACAGAAAAATGTCAAAAATCCACTGTCAACAAACCCACTACAAACAGTGAAC 4671 Db
QY
5119 AAGCTTTCCACACCTCCACAGGGGCGAGTGTGGACATCATCTCTGTAAAGAGGACAGAGC 5178 QY
Db
4672 AAGCTTTCCACACCTCCACAGGGGCGAGTGTGGACATCATCTCTGTAAAGAGGACAGAGC 4731 Db
QY
5179 AAAACCGTGGTCCACACGACAGTGAACAGTCCCTGACCAACCCAGGGAGGACACAGTGGT 5238 QY
Db
4732 AAAACCGTGGTCCACACGACAGTGAACAGTCCCTGACCAACCCAGGGAGGACACAGTGGT 4791 Db
QY
5239 ACATCTATGACTGTGAGCAAAAGAGTATTCACACAGAGCAAAAGTGAATGAAATTT 5298 QY
Db
4792 ACATCTATGACTGTGAGCAAAAGAGTATTCACACAGAGCAAAAGTGAATGAAATTT 4851 Db
QY
5299 TCAAGACCAAGAGAGTGTGAGTCAAGTCAAGTCTGACATCTGATAGAAAATTTGTTACC 5358 QY
Db
4852 TCAAGACCAAGAGAGTGTGAGTCAAGTCAAGTCTGACATCTGATAGAAAATTTGTTACC 4911 Db
QY
5359 AAGAGCACCAGAGAGCAATTTGTTTGGCTAATGATGACTTAAAAAAGTTGGCCCGA 5418 QY
Db
4912 AAGAGCACCAGAGAGCAATTTGTTTGGCTAATGATGACTTAAAAAAGTTGGCCCGA 4971 Db
QY
5419 AAAGGAGGAATCCGAGAGGTCCCTTATTTTAAATGCAATGCAAAACCTGCTTTGGATATA 5478 QY
Db
4972 AAAGGAGGAATCCGAGAGGTCCCTTATTTTAAATGCAATGCAAAACCTGCTTTGGATATA 5031 Db
QY
5479 TGGCCATATCTTCTAGACCGAGCTTTGGCATCTAGAGTATAGACTTCAGACA 5538 QY
Db
5032 TGGCCATATCTTCTAGACCGAGCTTTGGCATCTAGAGTATAGACTTCAGACA 5091 Db
QY
5539 GTAAAGTCTTACGTGAGTGAACCTGTGTTACGGTTACTCTGGGCAAGTTTGAGATGG 5598 QY
|||||

5092 GTAAAGTCTTACGTGAGTGAACCTGTATGTTTACGGTTACTCTGGGCAAGTTTGATGAGG 5151 Db
QY
5599 GATGATATGGCGCCCAAGGTTCTCTCAGGAGGAGGAGTACACGGACAGAAAATCCCGAA 5658 Db
Db
5152 GATGATATGGCGCCCAAGGTTCTCTCAGGAGGAGGAGTACACGGACAGAAAATCCCGAA 5211 Db
QY
5659 ACTGAAATCACAACACAGAAAATTAATAAGAGAGAGAGATGTTGGTCTTATGGCATTCGA 5718 QY
Db
5212 ACTGAAATCACAACACAGAAAATTAATAAGAGAGAGAGATGTTGGTCTTATGGCATTCGA 5271 Db
QY
5719 TTTGAAATATGATCAGGAAAATCATTTGTCATTTGGAGTTCCAGAAAACACAAAAGAA 5778 QY
Db
5272 TCTGAAATATGATCAGGAAAATCATTTGTCATTTGGAGTTCCAGAAAACACAAAAGAA 5331 Db
QY
5779 ACCCTTACACTCAGAGGAAAAGCCCTTCGATCAAGTGCATCGCGGCAAGAGACAGAA 5838 Db
Db
5332 ACCCTTACACTCAGAGGAAAAGCCCTTCGATCAAGTGCATCGCGGCAAGAGACAGAA 5391 Db
QY
5839 ACCCCCAAGCAAACTGGCCCTGTTTATTTATTTGAAACCTGGGTAGCAGAAAGAACTGGAA 5898 QY
Db
5392 ACCCCCAAGCAAACTGGCCCTGTTTATTTATTTGAAACCTGGGTAGCAGAAAGAACTGGAA 5451 Db
QY
5899 TTTGTTGGAGATCAGGCAATTTGCTGAGAGAGTGGAGAAAGAAAGCAAGCAGTTGAG 5958 Db
Db
5452 TTTGTTGGAGATCAGGCAATTTGCTGAGAGAGTGGAGAAAGAAAGCAAGCAGTTGAG 5511 Db
QY
5959 CAACAGGCTAAGAAACGACTGGAGCAGCAGAACCCGACAGTGAATGCAACTTCCACTACT 6018 QY
Db
5512 CAACAGGCTAAGAAACGACTGGAGCAGCAGAACCCGACAGTGAATGCAACTTCCACTACT 5571 Db
QY
6019 TCCCAACACAGCAGTACAACACGACCACTCTCCAGCAGACAGAAAGTTATGTTGGCCCC 6078 QY
Db
5572 TCCCAACACAGCAGTACAACACGACCACTCTCCAGCAGACAGAAAGTTATGTTGGCCCC 5631 Db
QY
6079 ATAAGTGGCTCAGTTACAACCTGGAACCAAAATGTTACTTAATGTTGATGTTGATGTTGAG 6138 QY
Db
5632 ATAAGTGGCTCAGTTACAACCTGGAACCAAAATGTTACTTAATGTTGATGTTGATGTTGAG 5691 Db
QY
6139 GCTACAGTACATTTCCAAACAAACAGAACTTTCACTAAACCTTTGCTACATGGTTAAG 6198 QY
Db
5692 GCTACAGTACATTTCCAAACAAACAGAACTTTCACTAAACCTTTGCTACATGGTTAAG 5751 Db
QY
6199 CAAGCCAGTCAAAATTCAGGCTGTTTCAAGTACAGCAGAAAGTCTCTGGTATCATTTCCA 6258 QY
Db
5752 CAAGCCAGTCAAAATTCAGGCTGTTTCAAGTACAGCAGAAAGTCTCTGGTATCATTTCCA 5811 Db
QY
6259 TCAAGTACAGGTACAGTACAGCAAACTTTACTTTTCAATTTCCAGCCAGGACAGCAAGTC 6318 QY
Db
5812 TCAAGTACAGGTACAGTACAGCAAACTTTACTTTTCAATTTCCAGCCAGGACAGCAAGTC 5871 Db
QY
6319 ACAATTTAGGCCCAATACCTCAGGCTCTGGAGGACCAACAGCAATTTCAAGTAAATCACA 6378 QY
Db
5872 ACAATTTAGGCCCAATACCTCAGGCTCTGGAGGACCAACAGCAATTTCAAGTAAATCACA 5931 Db
QY
6379 GGGCTCAGATTTCCGCTGTTATGACCGTGTATGAGAAACACACTTCCAAACAGTCAACACTA 6438 QY
Db
5932 GGGCTCAGATTTCCGCTGTTATGACCGTGTATGAGAAACACACTTCCAAACAGTCAACACTA 5991 Db
QY
6439 GGAAGGCAATTTATTCGAACACCTGTGATGTTAGCAGCCAGGTCCTCTCAGCAAGTGAATG 6498 QY
Db
5992 GGAAGGCAATTTATTCGAACACCTGTGATGTTAGCAGCCAGGTCCTCTCAGCAAGTGAATG 6051 Db
QY
6499 ACTCAATATCATCAGGGGCGCTGCTCTCCACTGCTCCAGTCTCCGCGCTTCAACACGGTTTCC 6558 QY
Db
6052 ACTCAATATCATCAGGGGCGCTGCTCTCCACTGCTCTCCAGTCTCCGCGCTTCAACACGGTTTCC 6111 Db
QY
6559 TCAACACTCTGGCAGAAAAGCTTAACTTTCAGAACTGCTCCACTTCAAAATATACAGTCTCA 6618 Db
6112 TCAACACTCTGGCAGAAAAGCTTAACTTTCAGAACTGCTCCACTTCAAAATATACAGTCTCA 6171 QY
6619 GCCTCACAACCCCTCGCCCGCCCAACAGGACAAGTGAAGTTCACCATGGCTCAACTTACT 6678 QY
6172 GCCTCACAACCCCTCGCCCGCCCAACAGGACAAGTGAAGTTCACCATGGCTCAACTTACT 6231 Db

QY	6679	CAGTTAAACACAGGGCCACGGTGGCAATCAAGGTTTGACAGTAGTAATTAAGGACAAGGT	6738	7759	AAGAAACAGCAACAGATAGAAATTAACGTGAACACACACCCCTCCAAGCTTCTTAATCAAGTT	7818
Db	6232	CAGTTAAACACAGGGCCACGGTGGCAATCAAGGTTTGACAGTAGTAATTAAGGACAAGGT	6291	6958	-----	6957
QY	6739	CAAACTACTGGACAGTTCAGTTGATACCTCAAGGGTGACGTACTCCCAGGCCAGGC	6798	7819	GAATCATTTACAAACACAGGTGCTGATGAAGCATAATGCTGTATAGAACATTTAAACACAG	7878
Db	6292	CAAACTACTGGACAGTTCAGTTGATACCTCAAGGGTGACGTACTCCCAGGCCAGGC	6351	6958	-----	7002
QY	6799	CAGCAGCTAATGCAAGCTGCAATGCAATGCTACTGTTACAGCGATTCCTCTTTACCCCA	6858	7879	AAAAAGAGCATGACTCCAGCTGAAAGAGAGAGAAATCAAAAGAAATGATTCTGTACACAG	7938
Db	6352	CAGCAGCTAATGCAAGCTGCAATGCAATGCTACTGTTACAGCGATTCCTCTTTACCCCA	6411	7003	AAAAAGAGCATGACTCCAGCTGAAAGAGAGAGAAATCAAAAGAAATGATTCTGTACACAG	7062
QY	6859	TTGGCAACACAGCCACAGCCAGCAGCCACCAACACACACCTGTTTCCACGACAGCAGCA	6918	7939	GTGATGAAGTATATTTTGGATAGATAGATAAGAGAAAGAAACAGCAGCAAAAAACGG	7998
Db	6412	TTGGCAACACAGCCACAGCCAGCAGCCACCAACACACACCTGTTTCCACGACAGCAGCA	6471	7063	GTGATGAAGTATATTTTGGATAGATAGATAAGAGAAAGAAACAGCAGCAAAAAACGG	7122
QY	6919	GCTACAGGTGAACAAGCGACAGTAAACTGTACACCCAGATCCAGGTACATCAACACAA	6978	7999	AAGCGTGAAGCAGTGTGGAGCAGAAACGTAGCAGCAGAAATGCCACTAAGCTGTACAGCT	8058
Db	6472	GCTACAGGTGAACAAGCGACAGTAAACTGTACACCCAGATCCAGGTACATCAACACAA	6531	7123	AAGCGTGAAGCAGTGTGGAGCAGAAACGTAGCAGCAGAAATGCCACTAAGCTGTACAGCT	7182
QY	6979	ACCCTGCCACAGCTCAGTCATCAAGTGTGGTCCAGCAAAAGCCAGCCAGCAGACTGCT	7038	8059	CTGCTCTTCAAGCACAACAGAGCAGCTCAGAGCCGAGATCCTGAAAGAGAGAGACACTCCCTG	8118
Db	6532	ACCCTGCCACAGCTCAGTCATCAAGTGTGGTCCAGCAAAAGCCAGCCAGCAGACTGCT	6591	7183	CTGCTCTTCAAGCACAACAGAGCAGCTCAGAGCCGAGATCCTGAAAGAGAGAGACACTCCCTG	7242
QY	7039	CAGCCTTCAGCTCGGCCAGCCCAACCCAGCCAGCCAGTCCCCAGCTCAGCCTGAAAGTT	7098	8119	GACAAAGATCTGCAAAATGAAGTGCAGGAAAGAGCTGAAGAGAGACTGAAAAATTAAGAAA	8178
Db	6592	CAGCCTTCAGCTCGGCCAGCCCAACCCAGCCAGCCAGTCCCCAGCTCAGCCTGAAAGTT	6651	7243	GACAAAGATCTGCAAAATGAAGTGCAGGAAAGAGCTGAAGAGAGACTGAAAAATTAAGAAA	7302
QY	7099	CAGACTCAGCCTGAAGTTTCAGACCCCAACCAACTGTTTCATCCCATGTCCCTTCTGAAGCA	7158	8179	GAAAAGACCTCATGCAGTTGGCTCAGGCCACAGCAGTAGCTGCACCCCTGCCCCCCAGTG	8238
Db	6652	CAGACTCAGCCTGAAGTTTCAGACCCCAACCAACTGTTTCATCCCATGTCCCTTCTGAAGCA	6711	7303	GAAAAGACCTCATGCAGTTGGCTCAGGCCACAGCAGTAGCTGCACCCCTGCCCCCCAGTG	7362
QY	7159	CAACCCACCCAGCAGTCATCCAAAGCCCAAGTTGCAGCAGCAGCTCAGCCTCAAAAGT	7218	8239	ACACCAGCTCTTCCAGCCCTCCAGCCCTCCACCTTCACTCCCTCCCTCCCTCCCTGCTG	8298
Db	6712	CAACCCACCCAGCAGTCATCCAAAGCCCAAGTTGCAGCAGCAGCTCAGCCTCAAAAGT	6771	7363	ACACCAGCTCTTCCAGCCCTCCAGCCCTCCACCTTCACTCCCTCCCTCCCTGCTG	7422
QY	7219	AATGTCCAAAGACAGTCTCCTGTTGTCGTTCCAAAGTCCATCAGACCTGGAATACGTCCA	7278	8299	CAACACACAGGCTTCTGTCCAGCCCACTTACCTGTGTTGCTCCAGAGAGAGAGAGCGG	8358
Db	6772	AATGTCCAAAGACAGTCTCCTGTTGTCGTTCCAAAGTCCATCAGACCTGGAATACGTCCA	6831	7423	CAACACACAGGCTTCTGTCCAGCCCACTTACCTGTGTTGCTCCAGAGAGAGAGAGCGG	7482
QY	7279	TCAACTCCATCCCACTGCTCCTGGACAACTCCAGGTTTCAGACTACACCTTCACAA	7338	8359	GAAGAGGAAAAGACTCCAGCTCAAGTCCAAAGAAAAGAAAATGATCTCTACTACTCA	8418
Db	6832	TCAACTCCATCCCACTGCTCCTGGACAACTCCAGGTTTCAGACTACACCTTCACAA	6891	7483	GAAGAGGAAAAGACTCCAGCTCAAGTCCAAAGAAAAGAAAATGATCTCTACTACTCA	7542
QY	7339	CGATTCCAAATTCACCCACATACATCTCTTCAGATACCTTCCCAAGGCCAGCCAGTCCA	7398	8419	AAGGAAACTAAGAAGGACACAAAGCTTACTGTATCTGTAAAACGCTTTATGATGAATCT	8478
Db	6892	CGATTCCAAATTCACCCACATACATCTCTTCAGATACCTTCCCAAGGCCAGCCAGTCCA	6951	7543	AAGGAAACTAAGAAGGACACAAAGCTTACTGTATCTGTAAAACGCTTTATGATGAATCT	7602
QY	7399	CAACCCAGGTCAGCTTCAACTCAACTCTTTTCATCAGGACAAACTTTAAATCAAGTT	7458	8479	A-----	8479
Db	6952	CAACCC-----	6957	7603	AAGTTCTATATTGGCTGTGATCTTTGTACTAACTGGTATCATGGAGAATGTGTGGCATC	7662
QY	7459	AGTGTTCATCCCATCCGCTCCTCAGCTACAAATACAGCAGCCACAGCCCAAGTCATT	7518	8480	-----	8479
Db	6958	-----	6957	7663	ACAGAAAAGGAGGCTAAGAAAATGGATGTACATCTGTAATGATTTAAACGGGCACAA	7722
QY	7519	GCTGTGCTCAGCTGCAACACAAAGTCCAGGTTCTCTCTCAGATCCAGTCAGAGTTGTG	7578	8480	-----	8484
Db	6958	-----	6957	7723	GAGGCAGCAGTGAGGAATTTGATCTATCTGCAGAACACCTTATGATGAGTCACAAATTT	7782
QY	7579	GCTCAGATACAGGCTCAGCAAGTGGTGTGCCCAAGCAAAATCAAACTCCAGTTACCTATC	7638	8485	TATATTGGCTGTGATCGGTGCAGAAATTTGGTACCATGGCGCTGGCTTGGCATCTTGCAA	8544
Db	6958	-----	6957	7783	TATATTGGCTGTGATCGGTGCAGAAATTTGGTACCATGGCGCTGGCTTGGCATCTTGCAA	7842
QY	7639	CAAAATTCAGCAAGCAGTCTGTGCAGACTCACCAGATTCAGAAATGCTGGTTACAGTCGAG	7698	8545	AGTGAGGACAGCTCATTGATGATGTCTGTCCACAGTGCAGTGCAGTGCAGAGGATGCC	8604
Db	6958	-----	6957	7843	AGTGAGGACAGCTCATTGATGATGTCTGTCCACAGTGCAGTGCAGTGCAGAGGATGCC	7902
QY	7699	GCAGCCAGTGTGCAAGAGCAGTTGCAAAAGGTTTCAGCAACTCAGGATCAGCAGCAAAAG	7758	8605	ATGACAGTCTCACGCCACTAACAGAGAGGATTTAGGGGTTGAAGAGGTTGCTCCCT	8664
Db	6958	-----	6957	7903	ATGACAGTCTCACGCCACTAACAGAGAGGATTTAGGGGTTGAAGAGGTTGCTCCCT	7962
				8665	TCCTTACAGGCCCAATAAGATGGCTTTCCTTTGAACCAAGTAGACCCCTAATGATGCA	8724

Db 7963 TCCTTACAGCCCATAGATGGCTTCCCTTGACACAGTAGACCCCTAATGATGA 8022
Qy 8725 CCAGATTATTATGTTTATTAGGAACCTATGGACCTTGCACCAATGGAAGAAGATGA 8784
Db 8023 CCAGATTATTATGTTTATTAGGAACCTATGGACCTTGCACCAATGGAAGAAGATGA 8082
Qy 8785 CAAGACGATATTATGAAGAAGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTGTAT 8844
Db 8083 CAAGACGATATTATGAAGAAGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTGTAT 8142
Qy 8845 AACTGTCGTTACTACAACTCACTCCCAATTTTACAGTGTGACAGAGTTCTCGAA 8904
Db 8143 AACTGTCGTTACTACAACTCACTCCCAATTTTACAGTGTGACAGAGTTCTCGAA 8202
Qy 8905 TCATCTTTTGTACAGAAATGAAGGCTTCAAGCTTACAGGTTCTATACACAACTG 8964
Db 8203 TCATCTTTTGTACAGAAATGAAGGCTTCAAGCTTACAGGTTCTATACACAACTG 8262
Qy 8965 CAGTCTACAGCTTCTTAA 8982
Db 8263 CAGTCTACAGCTTCTTAA 8280

RESULT 4

ABK84196
ID ABK84196 standard; cDNA; 2673 BP.
XX
AC ABK84196;
XX
DT 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #767.

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity.
XX
PS Claim 1; SEQ ID No 767; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)

CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, fungal infection and M5 is
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2673 BP; 887 A; 530 C; 678 G; 578 T; 0 other;

Query Match 26.0%; Score 2565.6; DB 24; Length 2673;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 2594; Conservative 0; Mismatches 29; Indels 1; Gaps 1;
Qy 216 GTACGATGACACGAGCGAGCGGTGGAGGAAGAGGAGGACATGCTCCGAGGAGCA 275
Db 31 GTACGGGGGGGGGGGGGGGGGGGGAGGAGGAAGAGGAGGACATGCTCCGAGGAGCA 90
Qy 276 GGAGGAGGAGGAGCGGCGACGCCGAGGAGAGCCAGGATTCGAGGACGAGGAGGATGA 335
Db 91 GGAGGAGGAGGAGCGGCGACGCCGAGGAGAGCCAGGATTCGAGGACGAGGAGGATGA 150
Qy 336 GATGGAAGAGGAGCGAGATGACTCCGATTCGCGAGGAGATGGAAGACGACGACGCA 395
Db 151 GATGGAAGAGGAGCGAGATGACTCCGATTCGCGAGGAGATGGAAGACGACGACGCA 210
Qy 396 CGCCAGTTACTGCACGGAAGAGCAGCTTCAGGAGCCATAGTACCTACAGCAGCACTCCAGG 455
Db 211 CGCCAGTTACTGCACGGAAGAGCAGCTTCAGGAGCCATAGTACCTACAGCAGCACTCCAGG 270
Qy 456 TAGCGGAAAACCAAGAGTACATCGGCTCTCTCTCTATATTTGGAGAAAACACATCCC 515
Db 271 TAGCGGAAAACCAAGAGTACATCGGCTCTCTCTCTATATTTGGAGAAAACACATCCC 330
Qy 516 GCCCTTTGAATTTCCCAAGTCCTGTAGGATTTAAATGGTGCCTAATGAGCATATAATGA 575
Db 331 GCCCTTTGAATTTCCCAAGTCCTGTAGGATTTAAATGGTGCCTAATGAGCATATAATGA 390
Qy 576 TGTCAATGCCATTTACGAGGTACTGCGGAACCTTTGGCAGCTGTTTGAGATATCTCCTTT 635
Db 391 TGTCAATGCCATTTACGAGGTACTGCGGAACCTTTGGCAGCTGTTTGAGATATCTCCTTT 450
Qy 636 TCGCTTTGAGGACTTTTGTGAGCTCTGTGAGCCAGAGCAGTGCACACTCATGGCAGA 695
Db 451 TCGCTTTGAGGACTTTTGTGAGCTCTGTGAGCCAGAGCAGTGCACACTCATGGCAGA 510
Qy 696 GATCGATGTTGTGCTTTTGAAGCAGTTCTCGGTGAAGAAGACACTTCCCAATACTACCTT 755
Db 511 GATCGATGTTGTGCTTTTGAAGCAGTTCTCGGTGAAGAAGACACTTCCCAATACTACCTT 570
Qy 756 TGGACCTGCTGATCTGAAGATAGCGTTAATTCACACTGTATTTTCATAGATGGGATGAC 815

Db	571	TGACCTGCTGATCTGAAGAATAGCGTTAAATTCACACTGTATTTTCATAGATGGGATGAC	630
QY	816	GTGGCCAGAGGTGCTGCGGGTCTACTGTGAGAGTGATAAGAGTAGTACCATCAGTTCTTCC	875
Db	631	GTGGCCAGAGGTGCTGCGGGTCTACTGTGAGAGTGATAAGAGTAGTACCATCAGTTCTTCC	690
QY	876	TTACCAAGAGGCGAGGAGTACCATATGAGACAGTAGAGAACAGATCAAAAGTTCTACA	935
Db	691	TTACCAAGAGGCGAGGAGTACCATATGAGACAGTAGAGAACAGATCAAAAGTTCTACA	750
QY	936	GTTTCTAGTCGATCAGTTCTTACACAAATATTTGCTCGAGAGGAATTCATGCTCGAAGG	995
Db	751	GTTTCTAGTCGATCAGTTCTTACACAAATATTTGCTCGAGAGGAATTCATGCTCGAAGG	810
QY	996	GGTGATACAGTATGATGACCATTTGAGGTTTGTGCACAACTTTGGGGATTTTCTTTGCTG	1055
Db	811	GGTGATACAGTATGATGACCATTTGAGGTTTGTGCACAACTTTGGGGATTTTCTTTGCTG	870
QY	1056	TGAGACATGTTTCAGCAGTATACCATTTTGGAAATGTTGTGAAGCCACCTCTTTGAGGAGTGCC	1115
Db	871	TGAGACATGTTTCAGCAGTATACCATTTTGGAAATGTTGTGAAGCCACCTCTTTGAGGAGTGCC	930
QY	1116	AGAGCAGGAGTGGCAGTGTGAAGTCTGTGTACACACACAGGTGCCTGGTGTGACTGACTG	1175
Db	931	AGAGCAGGAGTGGCAGTGTGAAGTCTGTGTACACACACAGGTGCCTGGTGTGACTGACTG	990
QY	1176	TCGTTGCTGAATCCCAAAAAATTAACCATATATTCGACATGAACCTATTGGATATATGATAG	1235
Db	991	TCGTTGCTGAATCCCAAAAAATTAACCATATATTCGACATGAACCTATTGGATATATGATAG	1050
QY	1236	AAGTCGGAGAAATACTGGTCTTGAACCGAAGACTCATAATAGAGAGATACAGAAAA	1295
Db	1051	AAGTCGGAGAAATACTGGTCTTGAACCGAAGACTCATAATAGAGAGATACAGAAAA	1110
QY	1296	TCAAAATGAAGAAAAATTTGGTATTACAGCACAAGGTCCTCAACTTGCAGAAATTAATGA	1355
Db	1111	TCAAAATGAAGAAAAATTTGGTATTACAGCACAAGGTCCTCAACTTGCAGAAATTAATGA	1170
QY	1356	CTGCTCTAGACAAAGATTTATGGGAAGCAGAACTCTGCAAAATTTCTAGAGAAATCGGTGA	1415
Db	1171	CTGCTCTAGACAAAGATTTATGGGAAGCAGAACTCTGCAAAATTTCTAGAGAAATCGGTGA	1230
QY	1416	AGAAATCCACCACACATGGACATTAACCTGAAGACCTGACCAATAGAGTTCGGGGCAGTAA	1475
Db	1231	AGAAATCCACCACACATGGACATTAACCTGAAGACCTGACCAATAGAGTTCGGGGCAGTAA	1290
QY	1476	CAAAATCCTTTCTGGCGCAGCTTAATCAAGAAATTTTGGAAATCCATAAGAGCCAAAAGGG	1535
Db	1291	CAAAATCCTTTCTGGCGCAGCTTAATCAAGAAATTTTGGAAATCCATAAGAGCCAAAAGGG	1350
QY	1536	AGACATTTGATAATGTTAAAGCCAGAGAAACAGAAAGAAAGAAATGAGACTGAGAA	1595
Db	1351	AGACATTTGATAATGTTAAAGCCAGAGAAACAGAAAGAAAGAAATGAGACTGAGAA	1410
QY	1596	TGACTCTAAAGATGCTGAGAAAAACAGAGAGAAATTTGAAGACCACTGCTCCCTTGAAGAAA	1655
Db	1411	TGACTCTAAAGATGCTGAGAAAAACAGAGAGAAATTTGAAGACCACTGCTCCCTTGAAGAAA	1470
QY	1656	CAGTCACGACAAACACACAGATGATCACCCTGAGCAAGGAAATCTGAGGAGCCACAGA	1715
Db	1471	CAGTCACGACAAACACACAGATGATCACCCTGAGCAAGGAAATCTGAGGAGCCACAGA	1530
QY	1716	AGTTGGGATAAGGTAACTCTGTGTGACCAATCTTGGCGACAACACAAACAAATGCAAC	1775
Db	1531	AGTTGGGATAAGGTAACTCTGTGTGACCAATCTTGGCGACAACACAAACAAATGCAAC	1590
QY	1776	TTTCAGAGAGACTAGTCCCTCTGAAGGGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCCC	1835
Db	1591	TTTCAGAGAGACTAGTCCCTCTCTGAAGGGAGGAGCCCTGTGGGGTGTCTCTCAGAAACCCC	1650
QY	1836	CGATAGCAGCACATGGCCAGAGAAAGAGTGGCATCTGAGTCCCCCAGGATGTGCCAGA	1895

RESULT 5
AAH15805
ID AAH15805 standard; cDNA; 2057 bp.
XX
AC AAH15805;
XX

Db	1651	CGATAGCAGCAACATGGCCAGAGAAAGGCTGGCATCTGAGCTCCCCCAGGATGTGCCAGA	1710
QY	1896	AGAACTTAAACAGACATGTGAGAGCAGTAACACTAGTGTACACTTACCTTCCATCCAGCC	1955
Db	1711	AGAACTTAAACAGACATGTGAGAGCAGTAACACTAGTGTACACTTCCATCCAGCC	1770
QY	1956	TAATCTGGAACACAGTAACAGCAGCAGTGAATAAATTTCTCCAGAGTGAATCTGCTAA	2015
Db	1771	TAATCTGGAACACAGTAACAGCAGCAGTGAATAAATTTCTCCAGAGTGAATCTGCTAA	1830
QY	2016	GGCAGCTGATGATCCTGAAATGGAGAAAGAACTCTACACCTGTCTCTATTTCAGGA	2075
Db	1831	GGCAGCTGATGATCCTGAAATGGAGAAAGAACTCTACACCTGTCTCTATTTCAGGA	1890
QY	2076	AGAGATAGTAGTGTATTTTTCACATCGGAGAAAGTCCACCGGGAGCTAAAGTGAATCTCCTGG	2135
Db	1891	AGAGATAGTAGTGTATTTTTCACATCGGAGAAAGTCCACCGGGAGCTAAAGTGAATCTCCTGG	1950
QY	2136	AGCTGAAAGAGAGCATCTGGCTCAACTCGAATCATCACAGATTCGCGAAATCCAGATAG	2195
Db	1951	AGCTGAAAGAGAGCATCTGGCTCAACTCGAATCATCACAGATTCGCGAAATCCAGATAG	2010
QY	2196	CAAACTTAGTCAAGTGAAGAGCCAGAGTGCAGCGCTGCACATGAAGCAAAATAAT	2255
Db	2011	CAAACTTAGTCAAGTGAAGAGCCAGAGTGCAGCGCTGCACATGAAGCAAAATAAT	2070
QY	2256	ATTTAAGGAGGCAAGAGGTACTGTTAGTTAACTCTCAAGGAGAAATTTACGGTTGAG	2315
Db	2071	ATTTAAGGAGGCAAGAGGTACTGTTAGTTAACTCTCAAGGAGAAATTTACGGTTGAG	2130
QY	2316	CACCAAAAAGAGTATCATGAAAGAAATATCAACAATATTTTAAATTTGGTCAAGA	2375
Db	2131	CACCAAAAAGAGTATCATGAAAGAAATATCAACAATATTTTAAATTTGGTCAAGA	2190
QY	2376	AGGGAAGTATCCGCTACCAACATCAATCTCCACCAATTCATTTGCTTTGAATAAGCA	2435
Db	2191	AGGGAAGTATCCGCTACCAACATCAATCTCCACCAATTCATTTGCTTTGAATAAGCA	2250
QY	2436	CCAGCACAGAGAGACCATGATAAGAGAGGATCTTGCACATAGTTCTGTCTGACTCC	2495
Db	2251	CCAGCACAGAGAGACCATGATAAGAGAGGATCTTGCACATAGTTCTGTCTGACTCC	2310
QY	2496	ACGAGAGAGTTCAATGGAAAGGTTCTGTCATGGTCCAAAGTTCTTACCATATCTAC	2555
Db	2311	ACGAGAGAGTTCAATGGAAAGGTTCTGTCATGGTCCAAAGTTCTTACCATATCTAC	2370
QY	2556	TCTGAGACTGACTATCACCCAATTTAGAAAACCAACATCCCTTCATCCTTTTTCATCCCAA	2615
Db	2371	TCTGAGACTGACTATCACCCAATTTAGAAAACCAACATCCCTTCATCCTTTTTCATCCCAA	2429
QY	2616	CTGGGCATCACATAGGGCAAAATTTGGATCAAGGCAGTTAGATGTGTAGCAAAACCCAGAGA	2675
Db	2430	CTGGGCATCACATAGGGCAAAATTTGGATCAAGGCAGTTAGATGTGTAGCAAAACCCAGAGA	2489
QY	2676	ATTTCGATTTGGCTTTAGCCATTTTGGAGTGTGAGTTAAACAGTTGTGTGATGCTACCAAT	2735
Db	2490	ATTTCGATTTGGCTTTAGCCATTTTGGAGTGTGAGTTAAACAGTTGTGTGATGCTACCAAT	2549
QY	2736	ATGGCGAGAAATTTTAGGGACATACCAGGTTACACCGGATGACATCAATTTGAAGAGAGA	2795
Db	2550	ATGGCGAGAAATTTTAGGGACATACCAGGTTACACCGGATGACATCAATTTGAAGAGAGA	2609
QY	2796	AAAGGAGAAAGTCAAAAAAAGAGAAAGAGGAAAGAGG 2839	
Db	2610	AAAGGAGAAAGTCAAAAAAAGAGAAAGAGGAAAGAGG 2653	

DT 26-JUN-2001 (first entry)
 XX Human cDNA sequence SEQ ID NO:14272.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 KW EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isozaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR Primer sets for synthesizing polynucleotides, particularly the 5602
 XX full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PT Claim 8; SEQ ID 14272; 2537pp + CD ROM; English.
 PS The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX Sequence 2057 BP; 566 A; 566 C; 435 G; 390 T; 0 other;
 SQ
 Query Match 17 48; Score 1719.4; DB 22; Length 2057;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1723; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Qy 5681 TAATTAAGAGGAGAGATGTTGGTCTTATGGCATTCGATTGTAATTTGATCAGGAAAA 5740
 Db 1 TAATTAAGAGGAGAGATGTTGGTCTTATGGCATTCGATTGTAATTTGATCAGGAAAA 60
 Qy 5741 TCATTTGTCCTATGGAGTTCCAGAACACACAAAGAAAGCCGTACACCTCAGAGGAAG 5800
 Db 61 TCATTTGTCCTATGGAGTTCCAGAACACACAAAGAAAGCCGTACACCTCAGAGGAAG 120
 Qy 5801 GCCTTCGATCAAGTGCACTCGGGCCAAAGAGACCCAGAAACGCCCAAGCAAACTGGCCCTG 5860
 Db 1201 CCAGCACCACCACCCACTGTTTCCACGACAGCAGCAGGTACAGGTGAAACAAGGCGAGA 1260

236 GTGAGAAATGGTACCATGGGCGCTGGCTTGGCATCTTGCAAAAGTGAGCGAGAGCTCATTTG 295
8564 ATGAGTATGCTCTGCACAGTGCAGTCAACAGAGATGCCATGACAGTGCCTCACGCCAC 8623
296 ATGAGTATGCTCTGCACAGTGCAGTCAACAGAGATGCCATGACAGTGCCTCACGCCAC 355
8624 TAACAGAGAAGGATTATGAGGGGTTGAAGAGGGTGTCTCGTTCCTTACAGGCCCATTAAGA 8683
356 TAACAGAGAAGGATTATGAGGGGTTGAAGAGGGTGTCTCGTTCCTTACAGGCCCATTAAGA 415
8684 TGGCCTGGCCCTTCCCTTGAACCAAGTAGACCCCTAATGATGCACAGATTAATATGGTGTGA 8743
416 TGGCCTGGCCCTTCCCTTGAACCAAGTAGACCCCTAATGATGCACAGATTAATATGGTGTGA 475
8744 TTAGGAACCTATGGACCTTGGCCACCATGGAAGAAAGAGTACAAAGACGATATATGAAA 8803
476 TTAGGAACCTATGGACCTTGGCCACCATGGAAGAAAGAGTACAAAGACGATATATGAAA 535
8804 AGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTTGATAACTGTGCTTACTCAATC 8863
536 AGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTTGATAACTGTGCTTACTCAATC 595
8864 CAAGTACTCCCAATTTTACAGTGTGCAGAACTTCTCGAATCATCTTTGTACAGAAAT 8923
596 CAAGTACTCCCAATTTTACAGTGTGCAGAACTTCTCGAATCATCTTTGTACAGAAAT 655
8924 TGAAGGCTTCAAAGCTAGCAGGTCTCATAAACAAACTGCGAGTCTACAGCTTCTTAAA 8983
656 TGAAGGCTTCAAAGCTAGCAGGTCTCATAAACAAACTGCGAGTCTACAGCTTCTTAAA 715
8984 GTTCAGCGTGTAACTTAAACATAAAACACAGCAAGAACTGGTGTGCTGAACATATTTTAA 9043
716 GTTCAGCGTGTAACTTAAACATAAAACACAGCAAGAACTGGTGTGCTGAACATATTTTAA 775
9044 ATTAAGAGCCAGATGTTTTAGTCAGGCTATCCTGACAGACTTGCACCTAAACTTCGTT 9103
776 ATTAAGAGCCAGATGTTTTAGTCAGGCTATCCTGACAGACTTGCACCTAAACTTCGTT 835
9104 TTTATTGTCTATAACAGTCCCAATATATTTCTTGCCCAATTTGTGCCAACGCAAGAAA 9163
836 TTTATTGTCTATAACAGTCCCAATATATTTCTTGCCCAATTTGTGCCAACGCAAGAAA 895
9164 AAGCAAGTCAACGACACCATATCTTGTCAAGATCAGATGGTTTTACTA-TTGTGCGAG 9222
896 AAGCAAGTCAACGACACCATATCTTGTCAAGATCAGATGGTTTTACTATTTGTGCGAG 955
9223 AAGCGAGAACTTTGTTTATT--GAAAAAAGAAAAAGAAAGCAAGAAAAAAGATA 9280
956 AAGCGAGAAATTTGTTTATTAATAAAGAAAAAGAAAGAAAGCAAGCAATGATA 1015
9281 CTATGGGTCAGTGTAACTCCATGGAAATGCCAGCTGCTCTTCAAGTGAAGAGCTGG 9340
1016 CTGTGGGTCAGTGTAACTCCATGGAAATGCCAGCTGCTCTTCAAGTGAAGAGCTGG 1075
9341 TTTAGAGTCTACAGAAACTTTTGACTGATTTATTTATTTGTTGCAAAAAGACGCTTT 9400
1076 TTTAGAGTCTACAGAAACTTTTGACTGATTTATTTATTTGTTGCAAAAAGAGCTTT 1135
9401 TTTATTGTGCGCTCATTTGTCAGTAACTAGTATTTTCTTATAAAATCCAGCCCGGTTA 9460
1136 TTTATTGTGCGCTCATTTGTCAGTAACTAGTATTTTCTTATAAAATCCAGCCCGGTTA 1195
9461 CATATAAT-CACTGTATCTTATC-ATGATTCCTGTAGGTAAAAGTACAGACACCTCT 9518
1196 CATATAATCCATCATATCTTATCAATGATTCCTGTAGTAAAAGTACAGACACCTCT 1255
9519 AGATCTCTTCTTCTATGAAGGAGCTGTATGTACACATGTGCACACACACAACT 9578
1256 AGATGCTCTTCTTCTATGAAGGAGCTGTATGTACACATGTGCACACACACAACT 1315
9579 GGGAAATCAACATGAGTTTATTTGTCATGGTAGATTAATAAGCTTGCATAAAGGTT 9638
1316 GCGAATCAACATGAGTTTATTTGTCATGGTAGATTAATAAGCTTGCATAAAGGTTG 1375

QY 9639 GGCTAAGTGGTCCCTTGGGCTACAGACTCTGTTGCCCTTGAATATACAGTACAAATTTGTCA 9698
Db ||||||||| ||| ||||||||| || ||||||||| || ||||||||| ||||||||| |||||
1376 GGCTAAGTGGTCC-TGGACTACAGACTCTGTTGCCCTTGAATATACAGTACAAATTTGTCA 1434
QY ||||||||| ||||||||| ||| ||||||||| || ||||||||| || ||||||||| ||||||||| |||||
9699 ATTACTCTCCACAGCTAAAGTGAGTAAATCTATTTTGAAGGTATCTTTGTTGTAACA 9758
Db ||||||||| ||||||||| ||| ||||||||| || ||||||||| || ||||||||| ||||||||| |||||
1435 ATTACTCCACAGCTTGAATGAGTAAATCTATTTGAAGGTATCTTTGTTGTAACA 1494
QY TTTGTGAGATCTAAATTTTTTTTCTTTTGTATTAAAAATTCACACTATGG 9805
Db ||||||||| ||||||||| ||| ||||||||| || ||||||||| || ||||||||| ||||||||| |||||
1495 TTGTGAGATCTAAATTTTTTCT-TTTGTATTAAAAATTCACACTATGG 1540

RESULT 7

AAAX30155

ID AAX30155 standard; DNA; 1032 BP.

XX AAX30155;

XX 18-JUN-1999 (first entry)

XX Human secreted protein gene 11.

XX DE

XX KW

KW

KW

KW

KW

KW

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

PR

PR

PR

PR

XX

PA

XX

PI

XX

DR

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Human; secreted protein; cancer; tumour; developmental abnormality;
foetal deficiency; blood disorder; immune system disorder; inflammation;
autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
digestive disorder; endocrine disorder; infection; AIDS; ss.

Homo sapiens.

WO9910363-A1.

04-MAR-1999.

27-AUG-1998;

98WO-US17709.

29-AUG-1997;

97US-0056271.

29-AUG-1997;

97US-0056073.

29-AUG-1997;

97US-0056247.

29-AUG-1997;

97US-0056270.

(HUMA-) HUMAN GENOME SCI INC.

Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;

WPI: 1999-190585/16.

P-PSDB; AAY04303.

New isolated human genes and the secreted polypeptides they encode

- useful for diagnosis and treatment of e.g. cancers, neurological

disorders, immune diseases, inflammation or blood disorders

Claim 1; Page 138; 170pp; English.

AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.

AAY04293 to AAY04321 represent the secreted proteins encoded by the 29

human genes. The genes and their corresponding secreted polypeptides are

useful for preventing, treating or ameliorating medical conditions, be

e.g. by protein or gene therapy. Also pathological conditions can be

diagnosed by determining the amount of the new polypeptides in a sample

or by determining the presence of mutations in the new genes. Specific

uses are described for each of the 29 genes, based on which tissues they

are most highly expressed in, and include developing products for the

diagnosis or treatment of cancer, tumours, developmental abnormalities

and foetal deficiencies, blood disorders, diseases of the immune system,

autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive

disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin

disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney

disorders, digestive/endocrine disorders, infections and AIDS. The

polypeptides are also useful for identifying their binding partners.

CC The sequences given in AAX30174 to AAX30182 and AAY04322 to AAY04334 are
CC used in the exemplification of the present invention.

XX
SQ Sequence 1032 BP; 343 A; 190 C; 227 G; 272 T; 0 other;
Query Match 8.3%; Score 814.6; DB 20; Length 1032;
Best Local Similarity 98.3%; Pred. No. 8.5e-152;
Matches 823; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 8444 TTTACTGTATCTGTAAGCCCTTATGATGAATCTAAATTTATATTTGCTGTGATCGGT 8503
Db 179 TGTACTGTATCTGCAACACCTTATGATGATGATCAATTTTATTTGCTGTGATCGGT 238
QY 8504 GTCAGAAATTGGTACCATGGGCGTGGTGGGATCTTGCAGAGTGAGGAGAGCTCATTTG 8563
Db 239 GTCAGAAATTGGTACCATGGGCGTGGTGGGATCTTGCAGAGTGAGGAGAGCTCATTTG 298
QY 8564 ATGAGTATCTGTCACAGTCCAGTCAACAGAGATGCCATGACAGTCTCAGCCAC 8623
Db 299 ATGAGTATCTGTCACAGTCCAGTCAACAGAGATGCCATGACAGTCTCAGCCAC 358
QY 8624 TAACAGAGAGGATTTATGAGGGTGTGAAGAGGGTGTCTCGTTCCTTACAGGCCATAAGA 8683
Db 359 TAACAGAGAGGATTTATGAGGGTGTGAAGAGGGTGTCTCGTTCCTTACAGGCCATAAGA 418
QY 8684 TGGCTGGCTTTCTTGAACAGTAGACCCCTTAATGATGACACAGATTTATTTGTTGA 8743
Db 419 TGGCTGGCTTTCTTGAACAGTAGACCCCTTAATGATGACACAGATTTATTTGTTGA 478
QY 8744 TTAAGGAACCTATGACCTTGGCCACCATGGAGAGAGATACAAAGACATATTATGAA 8803
Db 479 TTAAGGAACCTATGACCTTGGCCACCATGGAGAGAGATACAAAGACATATTATGAA 538
QY 8804 AGCTGACGAAATTTGTGGCAGATATGACCAAAATTTTGTACTGCTGTACTACAATC 8863
Db 539 AGCTGACGAAATTTGTGGCAGATATGACCAAAATTTTGTACTGCTGTACTACAATC 598
QY 8864 CAAGTGACTCCCATTTTACCAGTGTGCAAGAGTCTCTCGAATCATTTTGTACAGAAAT 8923
Db 599 CAAGTGACTCCCATTTTACCAGTGTGCAAGAGTCTCTCGAATCATTTTGTACAGAAAT 658
QY 8924 TGAAGGGCTTCAAGCTACAGGCTCTCAACAAACAACTGACGTACAGCTTCTTAA 8983
Db 659 TGAAGGGCTTCAAGCTACAGGCTCTCAACAAACAACTGACGTACAGCTTCTTAA 718
QY 8984 GTTCAGCGTGTAACTTAACATAAAACACAGCAAGAACTCTGTTGCTGAATATTATTA 9043
Db 719 GTTCAGCGTGTAACTTAACATAAAACACAGCAAGAACTCTGTTGCTGAATATTATTA 778
QY 9044 ATTAAGGAGCCAGATGTTTTAGTCAGGCTATCCTGACAAGACTTGACCTAAACTTCGTT 9103
Db 779 ATTAAGGAGCCAGATGTTTTAGTCAGGCTATCCTGACAAGACTTGACCTAAACTTCGTT 838
QY 9104 TTTATTGTCATAACAGTCCATTTATTTCTTGGCCAAATTTTGTCCACGGACAGAA 9163
Db 839 TTTATTGTCATAACAGTCCATTTATTTCTTGGCCAAATTTTGTCCACGGACAGAA 898
QY 9164 AAGCAAGTCAACGACACCATTTATCTGTCAAGATCAGATGTTTTTACTATTGTGGCAGA 9223
Db 899 AAGCAAGTCAACGACACCATTTATCTGTCAAGATCAGATGTTTTTACTATTGTGGCAGA 958
QY 9224 ACGGAGAAACCTTTGTTTATTGAAAAAAGAAAAAGAAAGAAAAAGATA 9280
Db 959 ACGGAGAAACCTTTGTTTATTGAAAAAAGAAAAAAGAAAAAAGAAAAA 1015

RESULT 8
AAH05331
ID AAH05331 standard; cDNA; 841 BP.
XX
AC AAH05331;
XX
DT 26-JUN-2001 (first entry)

XX
DE Human cDNA clone (5'-primer) SEQ ID NO:2166.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 2166; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 841 BP; 276 A; 204 C; 189 G; 169 T; 3 other;

Query Match 7.8%; Score 771.4; DB 22; Length 841;
Best Local Similarity 98.2%; Pred. No. 3e-143;
Matches 800; Conservative 0; Mismatches 13; Indels 2; Gaps 2;
QY 5682 AATTAAGAGGAGAGATGTTGGTCTTATGCGATTCGATTTGAATATTGTATCAGGAAAT 5741
Db 2 AATTAAGAGGAGAGATGTTGGTCTTATGCGATTCGATTCGATTTGAATATTGTATCAGGAAAT 61
QY 5742 CATTTTCTCCATTTGGAGTTCCAGAAACACAAAAGAACGCCCTACACCTCAGAGGAAAG 5801
Db 62 CATTTTCTCCATTTGGAGTTCCAGAAACACAAAAGAACGCCCTACACCTCAGAGGAAAG 121
QY 5802 CCTTCGATCAAGTGCACACTGCGCCAAAGAGACCAAGCCCAAGCAAACTGCCCTGT 5861
Db 122 CCTTCGATCAAGTGCACACTGCGCCAAAGAGACCAAGCCCAAGCAAACTGCCCTGT 181

KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX Homo sapiens.
OS WO200166719-A1.
XX 13-SEP-2001.
XX 02-MAR-2001; 2001WO-JP01629.
XX 07-MAR-2000; 2000JP-0159195.
XX (CHIB-) CHIBA PREFECTURE.
PA (HISM) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
XX WPI; 2001-565584/63.
DR Nucleic acids originating in gene expressed in human neuroblastoma,
XX useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents -
XX
PS Claim 1; Page 480; 2979pp; Japanese.
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
XX
XX Sequence 750 BP; 271 A; 108 C; 169 G; 183 T; 19 other;
SQ
Query Match 6.4%; Score 627.4; DB 22; Length 750;
Best Local Similarity 94.9%; Pred. No. 1.1e-114;
Matches 674; Conservative 0; Mismatches 32; Indels 4; Gaps 3;
QY 4005 TGACAGAGATGCCACCTCTGTCAAGAGCAATGGACTTTGAAGGAAACCTGGGATGTGA 4064
DB 42 TANCAGAGATCGGTTCTCTGTCAAGAGCAATGGACTTTGAAGGATTTCTGGGATGTGA 101
QY 4065 CTCTGAATCTAATAGCACTTTGGAAAATAGTCTGTATACCGTGTCTATTCAGATAGCAG 4124
DB 102 CTCTGAAGNACCCCACTTTGGAAAATAATCTGTATACCGTGTCTATTCAGATAGCAG 161
QY 4125 TGAAGAAGATATGATTGTTTCAGATAGCAATGAAGCAATTTCTGAACAGTTTCAGACTCG 4184
DB 162 TGAAGAAGATATGATTGTTTCAGATAGCAATGAAGCAATTTCTGAACAGTTTCAGACTCG 221
QY 4185 AGAACAGATGTTGAAGTCTTGGAGCCGTTAAAGTGTGAGTTGGTTCTGTGAGTCCAC 4244
DB 222 AGAACAGATGTTGAAGTCTTGGAGCCGTTAAAGTGTGAGTTGGTTCTGTGAGTCCAC 281
QY 4245 TGGAACTGTGAGGACAGGCTGCCGGTCAAGGGACTGAACAAATAGTGTAAACCAAG 4304
DB 282 TGGAACTGTGAGGACAGGCTGCCGGTCAAGGGACTGAACAAATAGTGTAAACCAAG 341
QY 4305 TCAGCAGAGAAATAGAGGAGAGACCAAGTTAAATAGTGTATCAATAAAGCTAAA 4364
DB 342 TCAGCAGAGAAATAGAGGAGAGACCAAGTTAAATAGTGTATCAATAAAGCTAAA 401
QY 4365 AAATACCACTGACAAAAAGATAATGAAATCGAGGTCTGAAAGAAAGGACAGAGAAC 4424
DB 402 AAATACCACTGACAAAAAGATAATGAAATCGAGGTCTGAAAGAAAGGACAGAGAAC 461
QY 4425 AAGTACATTCATAATGAAAGAGATAATAAACCCAAAATATATTTGAAAGGTGAATG 4484
DB 462 AAGTACATTCATAATGAAAGAGATAATAAACCCAAAATATATTTGAAAGGTGAATG 521
QY 4485 CTGTGAAAGAAATTTCTGAGAGTGTAGTGTAGTGTGTGTAATGTTGAACCAAGGTTAATAA 4544
|||||

Db 522 CTTGAAAGAAATTTCTGAGAGTAGTA-TAAGTGGTAATGTTGAACCAAGGTTAATAA 580
QY 4545 TATAAATAAAATATCCCTGAGATGATATTAATATCATTTGACTGTTAAAGAAATCTGCTAT 4604
|||||
Db 581 TATANATAAAATATCCCTGAGATGATATTAATATCATTTGACTGTTAAAGAAATCTGCTAT 640
QY 4605 AAGGCCATTCAATTAAGTGGTGTATCATCATGG--AGATTTTAAATGAAGAAACAGCTCC 4662
|||||
Db 641 AAGGCCATTCAATTAAGTGGTGTATCATCATGGGGAAGATTTTNTANTGAAGAAACAGCTCC 700
QY 4663 GAAACAAAATCGCATTTGCTGAGTCTTCTCAGATG-CTGAAGGTAACCTACC 4711
|||||
Db 701 GAAACAAAATCCCATTTGCGNGANTTCTCTCANATGCCCTGAAGGTAACCTACC 750
RESULT 12
ACA57390
ID ACA57390 standard; cDNA; 574 BP.
XX ACA57390;
XX 10-JUN-2003 (first entry)
DT XX
DE Human adipocyte Selected Interacting domain, SID, cDNA #477.
XX
XX Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
KW anorectic; antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
XX
XX Homo sapiens.
XX WO200286122-A2.
XX 31-OCT-2002.
XX
XX 14-MAR-2002; 2002WO-EP03768.
XX
XX 14-MAR-2001; 2001US-275734P.
XX (HYBR-) HYBRIGENICS.
PA Legrain P, Daviet L;
XX
XX WPI; 2003-103412/09.
DR P-PSDB; ABU70846.
XX
PT New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes -
XX
PS Claim 7; Page 267-268; 382pp; English.
XX
XX The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC recombinant host cell expressing a polypeptide in the adipocyte cells, a
CC polypeptide of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
CC sequences given in the specification (including its fragment or variant),
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC given in the specification (including its fragment or variant), a vector
CC comprising the SID (RTM) polynucleotide, a recombinant host cell
CC comprising the vector, a protein chip comprising the polypeptides and
CC a record comprising all or part of the data, listed in the specification.
CC The complex, polypeptides, polynucleotides and compounds are
CC useful for preventing or treating metabolic disorders such as obesity
CC or diabetes. The polynucleotides are useful as probes or primers. The
CC complex is particularly useful for identifying selected interacting

```
CC domains (SID (RTM)) for screening drugs that modulate the protein
CC interaction, thus exhibiting the therapeutic effect. The present
CC sequence encodes a SID (prey) protein of the invention.
XX
SQ Sequence 574 BP; 179 A; 123 C; 134 G; 138 T; 0 other;

Query Match          5.8%; Score 574; DB 25; Length 574;
Best Local Similarity 100.0%; Pred. No. 4e-104;
Matches 574; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8301 ACACACAGGCTTCTCTCCACGCCACCTTACCTGTGTCTCCAGAGAGAGGCGGA 8360
Db 1 ACACACAGGCTTCTCTCCACGCCACCTTACCTGTGTCTCCAGAGAGAGGCGGA 60
QY 8361 AGAGGAAAAGACTCCAGCTCAAGTCCCAAGAAAAGAAATGATCTCTACTACCTCAA 8420
Db 61 AGAGGAAAAGACTCCAGCTCAAGTCCCAAGAAAAGAAATGATCTCTACTACCTCAA 120
QY 8421 GGAACCTAAGAGGACACAAAGCTTTACTGTATCTCTAAACGCCCTTATGATGAATCTAA 8480
Db 121 GGAACCTAAGAGGACACAAAGCTTTACTGTATCTCTAAACGCCCTTATGATGAATCTAA 180
QY 8481 ATTTTATATTGGCTGTGATCGGTGTCAGAAATGGTACCATGGCGCTGCTTGGCATCTT 8540
Db 181 ATTTTATATTGGCTGTGATCGGTGTCAGAAATGGTACCATGGCGCTGCTTGGCATCTT 240
QY 8541 GCAAAGTGAGGACAGAGCTCATGTAGTATGCTGTCTCCACAGTGCACAGAGGA 8600
Db 241 GCAAAGTGAGGACAGAGCTCATGTAGTATGCTGTCTCCACAGTGCACAGAGGA 300
QY 8601 TGCATGACAGTGTCTACGCCACTACACAGAGGATTTATGAGGGTTGAAGAGGTGCT 8660
Db 301 TGCATGACAGTGTCTACGCCACTACACAGAGGATTTATGAGGGTTGAAGAGGTGCT 360
QY 8661 CCGTTCTTACAGGCCCATAAAGATGGCTTGGCTTTCTTGAACCACTAGACCTTAATGA 8720
Db 361 CCGTTCTTACAGGCCCATAAAGATGGCTTGGCTTTCTTGAACCACTAGACCTTAATGA 420
QY 8721 TGCACAGATATTATGGTGTATTAAAGAACCTATGGACCTTGGCCACCATGGAAGAAAG 8780
Db 421 TGCACAGATATTATGGTGTATTAAAGAACCTATGGACCTTGGCCACCATGGAAGAAAG 480
QY 8781 AGTACAAAGACGATATTATGAAAGCTGACGGAATTTGTGGCAGATATGACCAAAATTTT 8840
Db 481 AGTACAAAGACGATATTATGAAAGCTGACGGAATTTGTGGCAGATATGACCAAAATTTT 540
QY 8841 TGATAACTGTCGTTACTACAATCCAAAGTGACTCC 8874
Db 541 TGATAACTGTCGTTACTACAATCCAAAGTGACTCC 574

RESULT 13
AAH17323
ID AAH17323 standard; cDNA; 2123 BP.
XX
AC AAH17323;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:16734.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
```

```
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
PS Claim 8; SEQ ID 16734; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
SQ Sequence 2123 BP; 679 A; 439 C; 485 G; 520 T; 0 other;

Query Match          5.8%; Score 574; DB 22; Length 2123;
Best Local Similarity 99.1%; Pred. No. 6.7e-104;
Matches 577; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8664 TTCCTTACAGGCCCATAAAGATGGCTTGGCTTTCTTGAACCACTAGACCTTAATGTC 8723
Db 1542 TTCATTATAGGCCCATAAAGATGGCTTGGCTTTCTTGAACCACTAGACCTTAATGTC 1601
QY 8724 ACCAGATTATTATGGTGTATTAAAGAACCTATGGACCTTGGCCACCATGGAAGAAAGAGT 8783
Db 1602 ACCAGATTATTATGGTGTATTAAAGAACCTATGGACCTTGGCCACCATGGAAGAAAGAGT 1661
QY 8784 ACAAGACGATATTATGAAAGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTGA 8843
Db 1662 ACAAGACGATATTATGAAAGCTGACGGAATTTGTGGCAGATATGACCAAAATTTTGA 1721
QY 8844 TAACTCTGTTACTACAATCCCAAGTGTACTCCCATTTTACCAGTGTGTCAGAAATTTTTCGA 8903
Db 1722 TAACTCTGTTACTACAATCCCAAGTGTACTCCCATTTTACCAGTGTGTCAGAAATTTTTCGA 1781
QY 8904 ATCATTCTTTGTACAGAAATTTGAAAGCTTCAAAGCTTAGCAGGTCTCTATACCAACAACACT 8963
Db 1782 ATCATTCTTTGTACAGAAATTTGAAAGCTTCAAAGCTTAGCAGGTCTCTATACCAACAACACT 1841
QY 8964 GCAGTCTACAGCTTCTTAAAGTTCAGCGTGTAACTTAACATATAACACACAGCAAGATCT 9023
Db 1842 GCAGTCTACAGCTTCTTAAAGTTCAGCGTGTAACTTAACATATAACACACAGCAAGATCT 1901
QY 9024 GGTGTCTGAACCTATTTTTAAATTAAGAGGCCAGCATGTTTATTAGTCAGGCTATCCTGCAAA 9083
Db 9083 GGTGTCTGAACCTATTTTTAAATTAAGAGGCCAGCATGTTTATTAGTCAGGCTATCCTGCAAA 9083
```

Db	1902	GGTTGTCTGAACATATTTAAATTAAGSAGGCCAGATGTTTTTACTGAGGCTATCCTGACAA	1961
Qy	9084	GACTTCGACCTAAACTTCGTTTTTATTTGGTCATACACAGTCCAAATTATATTCTTGGCCAAATT	9143
Db	1962	GACTTCGACCTAAACTTCGTTTTTATTTGGCCATACACAGTCCAAATTATATTCTTGGCCAAATT	2021
Qy	9144	TTGTCCAACGGCAAGAAAAAGCAAGTCAACGACACCACTTATCTTGTCAAGATCAGAT	9203
Db	2022	TTGTCCACGGCACAGAAAAAGCAAGTCAACGACACCACTTATCTTGTCAAGATCAGAT	2081
Qy	9204	GGTTTTACTATTGTGCAGAACGAGAAAACTTTGTTTATTG	9245
Db	2082	GGTTTTACTATTGTGCAGAACGAGAAAACTTTGTTTATTG	2123

	RESULT 14	
	AAH09968/c	
	ID AAH09968 standard; cDNA; 556 BP.	
	XX	
	XX	
	AC AAH09968;	
	XX	
	XX	
	DT 26-JUN-2001 (first entry)	
	XX	
	XX	
	DE Human cdna clone (3'-primer) SEQ ID NO:6803.	
	XX	
	KW Human: primer: detection: diagnosis: antisense therapy: gene therapy: ss:	

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 3; SEQ ID 6803; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. RAH03166 to RAH13628 and

CC	AAH13633 to AAH18742 represent human cDNA sequences: AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.
XX	Sequence 556 BP; 158 A; 106 C; 112 G; 175 T; 5 other;
QQ	Query Match 5.5%; Score 538.2; DB 22; Length 556; Best Local Similarity 97.7%; Pred. No. 5.1e-97; Matches 543; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY	8690 GGCCTTTCCCTTGAACCACTAGACCTTAATGATGCACACAGATTAATATGGTCTTATTAAGG 8749
DB	
QY	8750 AACCTATGGACCTTGGCCACCATGGAAGAAAGAGTACAAAGACGATATATGAAAAAGCTGA 8809
DB	
QY	8810 CGGAATTGTGGCAGATATGACCAAAATTTTTGATAACTGTCGTACTACAAATCCAAAGT 8859
DB	
QY	436 CGGAATTGTGGCAGATATGACCAAAATTTTTGATAACTGTCGTACTACAAATCCAAAGC 377
DB	
QY	8870 ACTCCCACTTTACCACTGTGCAGAAAGTCTCGAATCATCTTTGTACAGAAATTTGAAG 8929
DB	
QY	8930 GCTTCAAGCTTAGCAGGTCTCATACACAACTGCAGTCTACAGCTCTTTAAAGTTCCAG 8989
DB	
QY	9050 GAGCAGATGTTTTAGTCAGGCTATCCTGACAGAGCTTGACCTAAAGCTCGTTTTTAT 9109
DB	
QY	9110 GGTCAATAACAGTCCCAATATATCTTGGCCAAATTTTGTCCAAAGGACAAAGAAAGCAA 9166
DB	
QY	9170 AGTCAACGACACCAATATCTTGTCAAGATCAGATGGTTTTACTATTTGGCAGAGCCAG 9229
DB	
QY	9230 AAACTTTGTTTATG 9245
DB	
QY	16 AAACTTTGTTTATG 1
DB	
RESULT 15	
AAZ17218	
ID	AAZ17218 standard; cDNA; 763 BP.
XX	AAZ17218;
AC	
DT	12-OCT-1999 (first entry)
DE	Human gene expression product cDNA sequence SEQ ID NO:4689.
XX	
KW	Human; gene; gene expression product; diagnosis; therapy; probe;
KW	detection; mapping; tissue typing; profiling; forensic; cancer;
KW	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9338972-A2.
XX	
PD	05-AUG-1999.
XX	
PF	28-JAN-1999; 99WO-US01619.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:12:14 ; Search time 12025 Seconds
(without alignments)
19938.766 Million cell updates/sec

Title: US-09-698-295-2
Perfect score: 9865
Sequence: 1 ggcaggctgagtgccgc.....acaaaaaaaaaaaaaaaaa 9865

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST.*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estin.*
 - 4: em_estmu.*
 - 5: em_estov.*
 - 6: em_estpl.*
 - 7: em_estro.*
 - 8: em_htc.*
 - 9: gb_est1.*
 - 10: gb_est2.*
 - 11: gb_htc.*
 - 12: gb_est3.*
 - 13: gb_est4.*
 - 14: gb_est5.*
 - 15: em_estfun.*
 - 16: em_estom.*
 - 17: em_gss_hum.*
 - 18: em_gss_inv.*
 - 19: em_gss_pln.*
 - 20: em_gss_vrt.*
 - 21: em_gss_fun.*
 - 22: em_gss_mam.*
 - 23: em_gss_mus.*
 - 24: em_gss_pro.*
 - 25: em_gss_rod.*
 - 26: em_gss_phg.*
 - 27: em_gss_vrl.*
 - 28: gb_gss1.*
 - 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	926	9.4	1086	12	BM458331
2	920.4	9.3	1131	12	BM802054
3	905.4	9.2	951	13	BU183985
4	892	9.0	1607	11	BC039881

5	857.8	8.7	897	13	BU187389
6	841	8.5	917	13	BQ716216
7	828	8.4	1046	12	BQ061478
8	823	8.3	867	13	BU169920
9	807.4	8.2	870	13	BQ718666
10	800.2	8.1	842	13	BU507358
11	799.4	8.1	883	13	BU176205
12	788.2	8.0	1200	11	AK082852
13	781.2	7.9	864	13	BQ717925
14	778	7.9	1050	12	BM451623
15	771.4	7.8	841	9	AU124130
16	732	7.4	1058	12	BM550180
17	727.2	7.4	786	13	BU739549
18	720.4	7.3	722	13	EX098782
19	713.8	7.2	905	10	BG164582
20	702.8	7.1	830	12	BI457269
21	693.6	7.0	711	14	CB305710
22	692.4	7.0	735	13	BQ575828
23	683.8	6.9	687	12	BQ021222
24	663.4	6.7	682	10	AW961200
25	662.4	6.7	683	2	HSM088021
26	661.4	6.7	746	12	BM675313
27	648.6	6.6	915	10	BG574526
28	638.2	6.5	968	13	BQ944857
29	636.8	6.5	648	12	BM312315
30	622.4	6.3	635	14	CB216388
31	620.4	6.3	894	14	CB181420
32	618.2	6.3	867	10	BE870101
33	615.2	6.2	695	10	BG393106
34	613	6.2	770	12	BI766147
35	612.4	6.2	673	12	BM992191
36	611	6.2	833	12	BI687006
37	609.8	6.2	643	13	BU674038
38	608.2	6.2	613	12	BQ001968
39	606	6.1	606	12	BM786451
40	601	6.1	677	9	AW390459
41	600.2	6.1	709	14	CB447515
42	595.2	6.0	741	9	AV704301
43	590.4	6.0	745	14	CB228541
44	589.8	6.0	718	12	BM996255
45	587.4	6.0	655	9	AV727211

ALIGNMENTS

RESULT 1
BM458331
LOCUS
DEFINITION AGNCOURT_6414182 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5496931
5', mRNA sequence.
ACCESSION BM458331
VERSION BM458331.1 GI:18507371
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1086)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: L1AM12126 row: d column: 20
High quality sequence stop: 702.

FEATURES		Location/Qualifiers	
source		1..1086	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:5496931"	
		/tissue_type="lymphoma, cell line"	
		/lab_host="DH10B (phage-resistant)"	
		/clone_lib="NIH_MGC_85"	
		[note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-df primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."]	
BASE COUNT		328 a	307 c 229 g 217 t 5 others
ORIGIN			
Query Match		9.4%;	Score 926; DB 12; Length 1086;
Best Local Similarity		97.5%;	Pred. No. 9.3e-118;
Matches 938;		Conservative 0;	Mismatches 24; Indels 0; Gaps 0;
QY	5959	CAACAGGCTAAGAAACGACTGGAGCAGCAGAGAGCGGACAGTGTGCAACTTCCACTACT	6018
Db	1	CAACAGGCTAAGAAACGACTGGAGCAGCAGAGAGCGGACAGTGTGCAACTTCCACTACT	60
QY	6019	TCCCAACAGCAGTACAAACAGCAGCACCATTCTCCAGCAGCAGAAAGTGTGCGGCCCC	6078
Db	61	TCCCAACAGCAGTACAAACAGCAGCACCATTCTCCAGCAGCAGAAAGTGTGCGGCCCC	120
QY	6079	ATAAGTGGCTCAGTTTACAACTGGAAACCAAAATGGTACTACTAAAGTTGGATCTCCA	6138
Db	121	ATAAGTGGCTCAGTTTACAACTGGAAACCAAAATGGTACTACTAAAGTTGGATCTCCA	180
QY	6139	GCTACAGTAACTTCCACAAACAGAACTTTCATCAAACTTGTCTACATGGTGAAG	6198
Db	181	GCTACAGTAACTTCCACAAACAGAACTTTCATCAAACTTGTCTACATGGTGAAG	240
QY	6199	CAAGGCCAGTCAAAATTCAGGCGTTTCAAGTACAGCAGAAAGTCTCGGGTATCATTTCA	6258
Db	241	CAAGGCCAGTCAAAATTCAGGCGTTTCAAGTACAGCAGAAAGTCTCGGGTATCATTTCA	300
QY	6259	TCAAGTACAGGTACAGTACGAAACCTTTACTTCCAGCCAGCAGCAGACATC	6318
Db	301	TCAAGTACAGGTACAGTACGAAACCTTTACTTCCAGCCAGCAGCAGACATC	360
QY	6319	ACAATTAGGCCCAATACCTCAGGCTCTGGAGAAACCAAGCAATTCAGATATATACA	6378
Db	361	ACAATTAGGCCCAATACCTCAGGCTCTGGAGAAACCAAGCAATTCAGATATATACA	420
QY	6379	GGGCTTCAGATTGGCCCTGGTATGACCGGTGATTAGAACACCACTCCACAGTCAACACTA	6438
Db	421	GGGCTTCAGATTGGCCCTGGTATGACCGGTGATTAGAACACCACTCCACAGTCAACACTA	480
QY	6439	GGAAAGCAATATTTCGAACACTGTGATGTGACGCCAGGTGCTCCTCAGCAAGTGATG	6498
Db	481	GGAAAGCAATATTTCGAACACTGTGATGTGACGCCAGGTGCTCCTCAGCAAGTGATG	540
QY	6499	ACTCAATATATCAGGGGGCAGCGCTGTCTCCACTGGCAGTCTCGGCCCTAACACGTTTCC	6558
Db	541	ACTCAATATATCAGGGGGCAGCGCTGTCTCCACTGGCAGTCTCGGCCCTAACACGTTTCC	600
QY	6559	TCAACACCTGGGAGAAAAGCTTAACTTCAGCAACGTTCACCTTCAATATACAGTCTTCA	6618
Db	601	TCAACACCTGGGAGAAAAGCTTAACTTCAGCAACGTTCACCTTCAATATACAGTCTTCA	660
QY	6619	GCCTCACACCCCTCGCCCAACAAAGGACAAAGTGAAGCTCACCAGTGCCTCAACTTACT	6678
Db	661	GCCTCACACCCCTCGCCCAACAAAGGACAAAGTGAAGCTCACCAGTGCCTCAACTTACT	720
QY	6679	CAGTTAACACAGGCCACGGTGGCAATCAAGGTTTGCAGAGTAGTAGTAATCAAGGACAAGT	6738
Db	721	CAGTTAACACAGGCCACGGTGGCAATCAAGGTTTGCAGAGTAGTAGTAATCAAGGACAAGT	780
QY	6739	CAAACTACTGGACAGTTGCAGTTGATACCTCAAGGGTGACTGTACTCCAGGCCCCAGGC	6798
Db	781	CAAACTACTGGACAGTTGCAGTTGATACCTCAAGGGTGACTGTACTCCAGGCCCCAGGC	840
QY	6799	CAGCAGCTAATGCAAGCTGCAATGCAAAATGGTACTGTTACGCGATTCTCTTTACCCCA	6858
Db	841	CAGCAGCTAATGCGAGCTGCAATGCGAAATGGTACTGGTCAAGGATCTCTTTACCCC	900
QY	6859	TTGCGAAACAGCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	6918
Db	901	ATTGCGACACAGCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCANCA	960
QY	6919	GG 6920	
Db	961	GG 962	
RESULT 2			
BM802054		1131 bp	mRNA linear EST 05-MAR-2002
LOCUS		ACENCOURT_6459616	NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5582055
DEFINITION		5', mRNA sequence.	
ACCESSION		BM802054	
VERSION		BM802054.1	GI:19118877
KEYWORDS		EST.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1 (bases 1 to 1131)	
AUTHORS		NIH-MGC http://mgc.nci.nih.gov/	
TITLE		National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL		Unpublished	
COMMENT		Contact: Robert Strausberg, Ph.D. Email: cgaabs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12342 row: o column: 16 High quality sequence stop: 670.	
FEATURES		source	
		1..1131	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:5582055"	
		/tissue_type="retinoblastoma"	
		/lab_host="DH10B (phage-resistant)"	
		/clone_lib="NIH_MGC_67"	
		[note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Library constructed by Life Technologies."]	
BASE COUNT		425 a	206 c 241 g 257 t 2 others
ORIGIN			
Query Match		9.3%;	Score 920.4; DB 12; Length 1131;
Best Local Similarity		91.4%;	Pred. No. 5.4e-117;
Matches 1019;		Conservative 0;	Mismatches 92; Indels 4; Gaps 4;
QY	2913	AGAGTACAGAGTGACAGGATATGGTGGTGGAGCTGAGTAACTCATGTTTATAG	2972
Db	12	AGAGTACAGAGTGACAGGATATGGTGGTGGAGCTGAGTAACTCATGTTTATAG	71
QY	2973	GTTCCTTCTTAATGCCAGGCAATACCTAATGTGAATACAGAAAGTCGTTAGAGGAAC	3032
Db	72	GTTCCTTCTTAATGCCAGGCAATACCTAATGTGAATACAGAAAGTCGTTAGAGGAAC	131
QY	3033	CAAAAATAATATGGATGAAATATGGATGAGTCAGTAAAGAAAATGTTCCAGGAATCC	3092

```
Db 132 CAAAAATATATGATGAAATATGGATGAGTCAGATAAAGAAATGTTCCAGAGTCC 191
QY 3093 AAAAAAATAAATAAGAGCTGATTTCTGAAAAAGATGAGTAAAAGTTTCAGATGCTGC 3152
Db 192 AAAAAAATAAATAAGAGCTGATTTCTGAAAAAGATGAGTAAAAGTTTCAGATGCTGC 251
QY 3153 AAAGAGGAGAGACCAAAATGAATGATATCTCAAGATTACTGAGAAGAGGACCAAGA 3212
Db 252 AAAGAGGAGAGACCAAAATGAATGATATCTCAAGATTACTGAGAAGAGGACCAAGA 311
QY 3213 TGTGAAGGAGCTCTTAGATTCTCAGAGTGAATAACCTGCAAGGAAGAACCAATGGAAGT 3272
Db 312 TGTGAAGGAGCTCTTAGATTCTCAGAGTGAATAACCTGCAAGGAAGAACCAATGGAAGT 371
QY 3273 AGACGATGACATGAAACAGAGTCACATGTAATTTGTCAGGAGAGTTCTCAAGTAGATGT 3332
Db 372 AGACGATGACATGAAACAGAGTCACATGTAATTTGTCAGGAGAGTTCTCAAGTAGATGT 431
QY 3333 GGTCAATGTTAGTCAGGGTTTTTCATCTAAGGACTAGTTACAAAAGAAACAAATCATC 3392
Db 432 GGTCAATGTTAGTCAGGGTTTTTCATCTAAGGACTAGTTACAAAAGAAACAAATCATC 491
QY 3393 CAAACTAGATGGACTTCTTCAAGAGGAGAAATTAACACAGTTTACACTGGAAGAAAACAGCG 3452
Db 492 CAAACTAGATGGACTTCTTCAAGAGGAGAAATTAACACAGTTTACACTGGAAGAAAACAGCG 551
QY 3453 ACTCGAAAAAATCAAGTTGGAGGGTGGAAATTAAGGGTATAGGAAAGACTTCTCAAAATTC 3512
Db 552 ACTCGAAAAAATCAAGTTGGAGGGTGGAAATTAAGGGTATAGGAAAGACTTCTCAAAATTC 611
QY 3513 TTCAAAAAATCTCTGTAATCACCAGTAATAAGCAAGCAAAAGAGGGTGCAGAGTGA 3572
Db 612 TTCAAAAAATCTCTGTAATCACCAGTAATAAGCAAGCAAAAGAGGGTGCAGAGTGA 671
QY 3573 CTCGATGAGACAAGACAGAGCCCAATGCAATATGATCAACCTGAGGACTTGATTCA 3632
Db 672 CTCGATGAGACAAGACAGAGCCCAATGCAATATGATCAACCTGAGGACTTGATTCA 731
QY 3633 GGGATGTTCAAAAGTGATTCCTCAGTCTTAGAATGAGTGATCTCCATCGAGGCCAGAACAA 3692
Db 732 GGGATGTTCAAAAGTGATTCCTCAGTCTTAGAATGAGTGATCTCCATCGAGGCCAGAACAA 791
QY 3693 CAAACTTATCCAAAGATCGAGTCTTAGATGATGTCCTCAGTCCAGGCCAGAACAA 3752
Db 792 ACACTTTATCC-AAAGATCAAGTGTAGATGATGTCCTCCTTCCGAGGCCAGAACAA 850
QY 3753 ATGTCGAAACAAATTTCCATTGAAATGACATAG-AAGAAAAAGTCTCTGACCTTGCCA 3811
Db 851 ATGTCGAAACAAATTTCCATTGAAATGACCTAGAAAAAAGTCCCTGACCTTGCCA 910
QY 3812 GTAGAGGCCAGGACCCACTTAAGAGTAATAACCAAAA-GGAATGATTTTTTTCATCGATGAC 3870
Db 911 GAAAGGCCAGGACCCGATCAGGAAAAACCAAGGAATTTGTTTCCCTCCATGGA 970
QY 3871 TCTAACTAGCCAGTCGAGATGATATTTGATCTTTCATCGATGATGTAAGAACAAAAACCGCTC 3930
Db 971 TTTCAACTAGCCGCGGCGAAGAAATTTGGAACCTTTTAAATCCCAAAAAAAGGACCGCTC 1030
QY 3931 ATACAGGAGGAAGTGACACCATTTGTTCTTCTTCCAGAGTCTTTACATTCATGAGT 3990
Db 1031 TTCACGAGAAAAATGGCCCTTTGTTCTTCTTCCAGAG-AGGTTTAAATTTCTCCAGG 1089
QY 3991 CCTAAAAGTACCAATGACAGAGATGCCACACCTCT 4025
Db 1090 CCCATCAGAACCTTGCCTAAAGGCCCCCTCCT 1124
```

RESULT 3

BU183985

LOCUS

DEFINITION

BU183985 951 bp mRNA linear EST 04-SEP-2002
ACENSCOURT_7971984 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6169696
5', mRNA sequence.

```
ACCESSION BU183985
VERSION BU183985.1 GI:22697969
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 951)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13535 row: d column: 17
High quality sequence stop: 631.
FEATURES
source
1..951
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6169696"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 376 a 169 c 191 g 215 t
ORIGIN
Query Match 9.28; Score 905.4; DB 13; Length 951;
Best Local Similarity 97.7%; Pred. No. 6.6e-115;
Matches 929; Conservative 0; Mismatches 21; Indels 1; Gaps 1;
QY 4190 AAGATGTTGAAGTCTTGGAGCCGCTTAAAGTGTGAGTTGGTTCTGGTGAGTCCACTGGAA 4249
Db 1 AAGATGTTGAAGTCTTGGAGCCGCTTAAAGTGTGAGTTGGTTCTGGTGAGTCCACTGGAA 60
QY 4250 ACTGTGAGGACAGCTCCCGTCAAGGGGACTGAAGCAAAATGGTAAAAAACAAGTCAGC 4309
Db 61 ACTGTGAGGACAGCTCCCGTCAAGGGGACTGAAGCAAAATGGTAAAAAACAAGTCAGC 120
QY 4310 AGAAGAAATTTAGAGGAGAGACCAGTTAAATAAATGATGATCAATAAAGCTAAAAATA 4369
Db 121 AGAAGAAATTTAGAGGAGAGACCAGTTAAATAAATGATGATCAATAAAGCTAAAAATA 180
QY 4370 CCACTGCAAAAAGAATAATGAAAATCGAGAGTCTGAAAAGAAGAGACAGACAAGTA 4429
Db 181 CCACTGCAAAAAGAATAATGAAAATCGAGAGTCTGAAAAGAAGAGACAGACAAGTA 240
QY 4430 CATTTCAAATTAATGGAAGAGATAATAACCCCAAAATATATTTCGAAAGGTGAATGCTTGA 4489
Db 241 CATTTCAAATTAATGGAAGAGATAATAACCCCAAAATATATTTCGAAAGGTGAATGCTTGA 300
QY 4490 AAGAAATTTCTGAGAGTAGAGTAGTAAGTGTGTAATTTGTAACCAAGGTTAATAATATAA 4549
Db 301 AAGAAATTTCTGAGAGTAGAGTAGTAAGTGTGTAATTTGTAACCAAGGTTAATAATATAA 360
QY 4550 ATAAATAATCCCTGAGAATGATTAATAATCATTGACTGTTAAAGAATCTGCTATAAGGC 4609
Db 361 ATAAATAATCCCTGAGAATGATTAATAATCATTGACTGTTAAAGAATCTGCTATAAGGC 420
QY 4610 CATTCAATTAATGTTGATGTCATCATGCAAGATTTTAATGAAGAAACAGCTCCCAACAA 4669
Db 421 CATTCAATTAATGTTGATGTCATCATGCAAGATTTTAATGAAGAAACAGCTCCCAACAA 480
```

QY 4670 AATGCCATTTGCTGAGTTCTTACAGTCTGAAGGTAATACTACCGAGATAGCCTTGAGACCC 4729
|||||
Db 481 AATGCCATTTGCTGAGTTCTTACAGTCTGAAGGTAATACTACCGAGATAGCCTTGAGACCC 540
|||||
QY 4730 TGCCATCAACAAAGAGCTGACAGTACACAGAGACCCACACCTCAGCATCTTGTCCAG 4789
|||||
Db 541 TGCCATCAACAAAGAGCTGACAGTACACAGAGACCCACACCTCAGCATCTTGTCCAG 600
|||||
QY 4790 AAAGCAATTCAGTTAATCAGGTAGAGATGGAATAGAACTCAGAGTAAAGAAAG 4849
|||||
Db 601 AAAGCAATTCAGTTAATCAGGTAGAGATGGAATAGAACTCAGAGTAAAGAAAG 660
|||||
QY 4850 TTACTTCATCATCTTACTTCTTCAAGAGGAATCTAATCTCAGTAATGACTTTATTGATG 4909
|||||
Db 661 TTACTTCATCATCTTACTTCTTCAAGAGGAATCTAATCTCAGTAATGACTTTATTGATG 720
|||||
QY 4910 AAAATGGTCTCCCATCAACAAAATGAATGTCAATGGAGATCTAAAGAAACCG 4969
|||||
Db 721 AAAATGGTCTCCCATCAACAAAATGAATGTCCATGGAGATCTAAAGAAACCG 780
|||||
QY 4970 TCATCACAAGAGTCAACAGATGACCTCCACAGTGGCCACAGATCAAAATCTGTATCA 5029
|||||
Db 781 TCATCACAAGAGTCAACAGATGACCTCCACAGTGGCCACAGATCAAAATCTGTATCA 840
|||||
QY 5030 AGGTAGAAAAGGGGATAAGCAACTGT-GGTTTCTTCCACAGAAAATTTGCAAAATCC 5088
|||||
Db 841 AGGTAGAAAAGGGGATAAGCAACTGTGGGTTTCTTCCACAGAAAATTTGCAAAATCC 900
|||||
QY 5089 ACTGTCAACACACCTACAAAGTACCAAGCTTTTCCACACCCCTCCACA 5139
|||||
Db 901 ACTGTCCCAACACCTACCAAGTACCAAGCTTTTCCCAACCCCTCAA 951
|||||

RESULT 4
BC039881
LOCUS BC039881 1607 bp mRNA linear HTC 04-MAR-2003
DEFINITION Homo sapiens, similar to p300/CBP-associated factor, clone
IMAGE:5113116, mRNA.
ACCESSION BC039881
VERSION BC039881.1 GI:25058861
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1607)
Strausberg, R.
Direct Submission
Submitted (15-NOV-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhgri.nih.gov
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 78 Row: d Column: 19
This clone has the following problem: retained intron.

FEATURES
source
1..1607
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5113116"
/tissue_type="Cervix, carcinoma"
/clone_lib="NIH_MGC_12"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
BASE COUNT 531 a 319 c 356 g 401 t
ORIGIN

Query Match 9.08; Score 892; DB 11; Length 1607;
Best Local Similarity 85.8%; Pred. No. 4e-113;
Matches 1079; Conservative 0; Mismatches 5; Indels 174; Gaps 1;
QY 7864 GAACATTTAAACAGAAAAAGAGCATGACTCCAGCTGAAAGAGAGAGAAATCAAGAATG 7923
Db 12 GAACATTTAAACAGAAAAAGAGCATGACTCCAGCTGAAAGAGAGAGAAATCAAGAATG 71
QY 7924 ATTCTCTGTAAACAGGTGATGAGTATATTTTGGATAGATAGATATAAGAGAAAAACAG 7983
Db 72 ATTCTCTGTAAACAGGTGATGAGTATATTTTGGATAGATAGATATAAGAGAAAAACAG 131
QY 7984 GCACAAAAAAACGGAAGCGTGAAGAGAGTGTGAGCAGAGAACTAGCAGCAGAAATGCC 8043
Db 132 GCACAAAAAAACGGAAGCGTGAAGAGAGTGTGAGCAGAGAACTAGCAGCAGAAATGCC 191
QY 8044 ACTAAGCTGTGAGTCTGCTCTTCAAGCACAAGAGCAGCTCAGAGCCGAGATCCTGAAG 8103
Db 192 ACTAAGCTGTGAGTCTGCTCTTCAAGCACAAGAGCAGCTCAGAGCCGAGATCCTGAAG 251
QY 8104 AAGAGAGCACTCTGGACAGAGTCTGCAATTTGAAGTGCAGGAGAGAGCTGAAGAGAGAC 8163
Db 252 AAGAGAGCACTCTCTGGACAGAGTCTGCAATTTGAAGTGCAGGAGAGAGCTGAAGAGAGAC 311
QY 8164 CTGAAAAATTAAGAAAAAGAAAGACCTGATGAGTTGGCTCAGGCCACAGCAGTAGTGTCA 8223
Db 312 CTGAAAAATTAAGAAAAAGAAAGACCTGATGAGTTGGCTCAGGCCACAGCAGTAGTGTCA 371
QY 8224 CCCTGCCCCCGAGTGACACAGTCTTCCAGCCCTCAGGCCCTCAGCTTACCTTCCCTCC 8283
Db 372 CCCTGCCCCCGAGTGACACAGTCTTCCAGCCCTCAGGCCCTCAGCTTACCTTCCCTCC 431
QY 8284 CCTCCAGCTGTGTGACACACAGGCTTCTGTCCAGCCCTTACCTTGTGTGTCTTCC 8343
Db 432 CCTCCAGCTGTGTGACACACAGGCTTCTGTCCAGCCCTTACCTTGTGTGTCTTCC 491
QY 8344 CAGAAGAGGAAAGCGGAAGAGAAAAAGACTCCAGCTCAAAGTCCAGAAAAAGAAATG 8403
Db 492 CAGAAGAGGAAAGCGGAAGAGAAAAAGACTCCAGCTCAAAGTCCAGAAAAAGAAATG 551
QY 8404 ATCTCTACTACTCTCAAGGAAACTAAGAGAGACACAAAGCTTTTACTGTATCTGTAAACG 8463
Db 552 ATCTCTACTACTCTCAAGGAAACTAAGAGAGACACAAAGCTTTTACTGTATCTGTAAACG 611
QY 8464 CCTTATGATGATCTA----- 8479
Db 612 CCTTATGATGATCTAAGTTCTATATTTGGCTGTGATCTTTTGTACTTAAGTCTGATCATGGA 671
QY 8480 ----- 8479
Db 672 GAATGTGTGGATCACAGAAAAAGGAGGCTAAGAAAAATGGATGTGTATCATCTGTAATGAT 731
QY 8480 ----- 8479
Db 732 TGTAACGGGCACAGAGGGGAGCAGTGAAGAAATGTGTACTGTATCTGCAGAACACCTTAT 791


```

ACCESSION B0716216
VERSION B0716216.1 GI:21855113
KEYWORDS EST. Homo sapiens (human)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: ccapbs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13592 row: d column: 21
High quality sequence stop: 680.
FEATURES
source
Location/Qualifiers
1..917
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6191588"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_lib="Lupski-sympathetic_trunk"
/clone_vector="pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; CDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGGCTCCG-3' and
5'-GACTAGTTCTAGATCGGAGCGGCCGCTT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."
BASE COUNT 284 a 251 c 196 g 176 t 10 others
ORIGIN
Query Match: 8.5%; Score 841; DB 13; Length 917;
Best Local Similarity 98.5%; Pred. No. 4.8e-106;
Matches 867; Conservative 0; Mismatches 10; Indels 3; Gaps 2;
QY 5809 TCAAGTGCACCTGCGCCCAAGAGACACAGAACGCCCAAGCAACTGGCCCTGTATTATT 5868
DB 1 TCAAGTGCACCTGCGCCCAAGAGACACAGAACGCCCAAGCAACTGGCCCTGTATTATT 60
QY 5869 GAAACCTGGGTAGCAGAGAGAACTGGAATTTGGGAGATCAGGGCATTTGCTGAGAGA 5928
DB 61 GAAACCTGGGTAGCAGAGAGAACTGGAATTTGGGAGATCAGGGCATTTGCTGAGAGA 120
QY 5929 GTGGAGAAAGAAAGGCAACAGCAGTTGAGCAACAGGCTAAGAAACGACTGGAGCAGCAG 5988
DB 121 GTGGAGAAAGAAAGGCAACAGCAGTTGAGCAACAGGCTAAGAAACGACTGGAGCAGCAG 180
QY 5989 AAGCGGACAGTGATTTGCAACTTCCACTACTTCCCAACAAGCAGTACACCAAGCAGCATC 6048
DB 181 AAGCGGACAGTGATTTGCAACTTCCACTACTTCCCAACAAGCAGTACACCAAGCAGCATC 240
QY 6049 TCTCCAGCAGAGAAAGTTATGGTGGCCCCCATAAGTGGCTCAGTTACAACTGGAAACAAA 6108
DB 241 TCTCCAGCAGAGAAAGTTATGGTGGCCCCCATAAGTGGCTCAGTTACAACTGGAAACAAA 300
QY 6109 ATGGTACTAACTACTAAAGTTGGATCTCCAGCTCAGTAACTTCCCAACAAAGAAAC 6168
DB 301 ATGGTACTAACTAAAGTTGGATCTCCAGCTCAGTAACTTCCCAACAAAGAAAC 360

```

```

QY 6169 TTTTCATCAAAACCTTTTCTACATGGGTTAAGCAAGGCCAGTCAAATTCAGGCGTTGTCAA 6228
DB 361 TTTTCATCAAAACCTTTTCTACATGGGTTAAGCAAGGCCAGTCAAATTCAGGCGTTGTCAA 420
QY 6229 GTACAGCAGAAAGTCTCTGGGTATCATTCATCAAGTACAGGTACAGTACAGCAAACTTTT 6288
DB 421 GTACAGCAGAAAGTCTCTGGGTATCATTCATCAAGTACAGGTACAGTACAGCAAACTTTT 480
QY 6289 ACTTCATTCAGCCAGGACAGCAACAGTCAACAATTAGCCCAATACCTCAGGCTCTGGA 6348
DB 481 ACTTCATTCAGCCAGGACAGCAACAGTCAACAATTAGCCCAATACCTCAGGCTCTGGA 540
QY 6349 GGAACCCACAAGCAATTCACAAGTAATCACAGGSCCTCAGATTCGCCCTGGTATGACCGTG 6408
DB 541 GGAACCCACAAGCAATTCACAAGTAATCACAGGSCCTCAGATTCGCCCTGGTATGACCGTG 600
QY 6409 ATTAGAACCACTCTCAACAGTCAACACTAGGAAAGCAATTTATTCGAACACTGTGATG 6468
DB 601 ATTAGAACCACTCTCAACAGTCAACACTAGGAAAGCAATTTATTCGAACACTGTGATG 660
QY 6469 GTACAGCAGGTGCTCTCTCAGCAAGTGATGACTCAAAATCATCAGGGGGCAGGCTGTCTCC 6528
DB 661 GTACAGCAGGTGCTCTCTCAGCAAGTGATGACTCAAAATCATCAGGGGGCAGGCTGTCTCC 720
QY 6529 ACTGCAGTCTCCGCCCTTAACAGGGTTTCTCAACACTGGGGCAGGAGAAAGCTTAACCTCA 6588
DB 721 ACTGCAGTCTCCGCCCTTAACAGGGTTTCTCCCTCACACCTGNGCAGAAAGCTTAACCTCA 780
QY 6589 GCAAGCTCCACTTCAATATACAGTCTTCAGCCTCAACACCTGGGGCAGGAGAAAGCTTAACCTCA 6648
DB 781 GCAAGCTCCACTTCAATATACAGTCTTCAGCCTCAACACCTGGGGCAGGAGAAAGCTTAACCTCA 840
QY 6649 C-AAGTGAAGCTCACCATGG--CTCAACTTACTCAGTTAA 6685
DB 841 C-AAGTGAAGCTCACCATGGGCTCAGTTAACTTACTCAGTTAA 880

```

```

RESULT 7
BQ061478
LOCUS BQ061478
DEFINITION BQ061478 NIH_MGC_99 Homo sapiens cdna clone IMAGE:5920928
5', mRNA sequence.
ACCESSION BQ061478
VERSION BQ061478.1 GI:19885571
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1046)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccapbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM2085 row: g column: 09
High quality sequence stop: 674.
FEATURES
source
Location/Qualifiers
1..1046
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5920928"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"

```


/clone_lib="NIH_MGC_99"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

```
BASE COUNT      368 a  226 c  209 g  243 t
ORIGIN
Query Match      8.4%; Score 828; DB 12; Length 1046;
Best Local Similarity 96.9%; Pred. No. 2.9e-104;
Matches 899; Conservative 0; Mismatches 20; Indels 9; Gaps 5;
QY 4507 AGAGTAGTAAGTGGTAATGTTGAACCAAGGTTAATAATATAATAATAATAATATCCCTGAG 4566
DB 1 AGAGTAGTAAGTGGTAATGTTGAACCAAGGTTAATAATATAATAATAATAATATCCCTGAG 60
QY 4567 AATGATATTAAATCATTCGACTGCTTTAAAGAAATCGCTATAAGGCCATTCAATTAATGGTGT 4626
DB 61 AATGATATTAAATCATTCGACTGCTTTAAAGAAATCGCTATAAGGCCATTCAATTAATGGTGT 120
QY 4627 GTCATCATGGAAGATTTTAATGAAGAAACACAGCTCCGAAACAAATATCGCATTTGCTGAGT 4686
DB 121 CTCATCATGGAAGATTTTAATGAAGAAACACAGCTCCGAAACAAATATCGCATTTGCTGAGT 180
QY 4687 TCTTCAGATGCTGAAGTAACTACCGAGATACGCTTGAGACCCCTGCCATCAACCAAGAG 4746
DB 181 TCTTCAGATGCTGAAGTAACTACCGAGATACGCTTGAGACCCCTGCCATCAACCAAGAG 240
QY 4747 TCTGACGTACACAGACGACACACCCCTCAGCATCTTTGCCAGAAAGCAATTCAGTTAAT 4806
DB 241 TCTGACGTACACAGACGACACACCCCTCAGCATCTTTGCCAGAAAGCAATTCAGTTAAT 300
QY 4807 CAGGTAGAAGATATGGAATAGAAACCTCAGAAGTTAAGAAGTTACTTCATCACCCTATT 4866
DB 301 CAGGTAGAAGATATGGAATAGAAACCTCAGAAGTTAAGAAGTTACTTCATCACCCTATT 360
QY 4867 ACTTCTGAAGGAATCTAATCTCAGTATGATGACTTTATTGATGAATGGTCTGCCCATC 4926
DB 361 ACTTCTGAAGGAATCTAATCTCAGTATGATGACTTTATTGATGAATGGTCTGCCCATC 420
QY 4927 AACAAAAATGAAATGTCAATGGAGAATCTTAAAGAAAAACCGTCAATCAGCAAGTCAACC 4986
DB 421 AACAAAAATGAAATGTCAATGGAGAATCTTAAAGAAAAACCGTCAATCAGCAAGTCAACC 480
QY 4987 ACGATGACCTCACAGTGGCCACAGAATCAAAAATGTGATCAAGGTAGAAAAAGGCGAT 5046
DB 481 ACGATGACCTCACAGTGGCCACAGAATCAAAAATGTGATCAAGGTAGAAAAAGGCGAT 540
QY 5047 AAGCAAACTGTGGTTCTTCCACAGAAAATGTGCAGAAATCCACTGTCCACACCACTACT 5106
DB 541 AAGCAAACTGTGGTTCTTCCACAGAAAATGTGCAGAAATCCACTGTCCACACCACTACT 600
QY 5107 ACAACAGTGACCAAGCTTTCCACACCCCTCCACAGCGCGAGTGTGGACATCATCTCTGTA 5166
DB 601 ACAACAGTGACCAAGCTTTCCACACCCCTCCACAGCGCGAGTGTGGACATCATCTCTGTA 660
QY 5167 AAGGAGCAGAGAAACCGTGTCCACAGCAGTGCAGACTTCCCTGACCAACACGCGGA 5226
DB 661 AAGGAGCAGAGAAACCGTGTCCACAGCAGTGCAGACTTCCCTGACCAACACGCGGA 720
QY 5227 GGCACACTGGTTACATCTATGACTGTGAGCAAGAGTATCCACAGCAGAGCAAAAGTGAAA 5286
DB 721 GGCACACTGGTTACATCTATGACTGTGAGCAAGAGTATCCACAGCAGAGCAAAAGTGAAA 780
QY 5287 CTGATGAATTTTCAAGA-CCAAAAGAGACT-CGTTTCAGGTACAGCTCTGCCAT-CCTAT 5343
DB 781 CTGATGAATTTTCAAGACCCCAAGAAAGACTCCGTTTCAGGTACAGCTCTGCCATCCCTAT 840
```

```
QY 5344 AGAAAAATTTGTACCAAG-----AGCACCAAGAAAGACCATTTTGTTCCTCTAA--TGAT 5397
DB 841 AGAAAAATTTGTACCAAGGAGCAGCAAGAAAGAGACGATTTTGTTCCTCTAAATGATG 900
QY 5398 GACTTAAAAAAGTTGGCCGCGAAAGAGAG 5425
DB 901 GACTTAAAAAAGTTGGCCGCGAAAGAGG 928

RESULT 8
BU169920'
LOCUS      BU169920          867 bp      mRNA      linear      EST 04-SEP-2002
DEFINITION AGENCOURT_7938595 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6012389
5', mRNA sequence.
ACCESSION  BU169920
VERSION    BU169920.1 GI:22683904
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 867)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM13203 row: j column: 06
          High quality sequence stop: 690.
          Location/Qualifiers
            1..867
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:6012389"
              /tissue_type="embryonal carcinoma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_92"
              /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
              Average insert size 2.5 kb. Library enriched for
              full-length clones and constructed by Life Technologies.
              Note: this is a NIH_MGC Library."
BASE COUNT  303 a  170 c  195 g  196 t  3 others
ORIGIN
Query Match      8.3%; Score 823; DB 13; Length 867;
Best Local Similarity 97.8%; Pred. No. 1.5e-103;
Matches 843; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
QY 1970 GTAACAGCAGCAGTGAACCTAAATTTCTCCACAGTGAATCTCTAAGGCAGCTGATGTC 2029
DB 1 GTAACAGCAGCAGTGAACCTAAATTTCTCCACAGTGAATCTCTAAGGCAGCTGATGTC 60
QY 2030 CTGAAAAATGGGAAAGAGAAATCTATACACCTGTCTCTATTTCAGGAAGAGATAGTAGGTG 2089
DB 61 CTGAAAAATGGGAAAGAGAAATCTATACACCTGTCTCTATTTCAGGAAGAGATAGTAGGTG 120
QY 2090 ATTTTCATCATCGAGAGATCCACCGGGAGCTAAAGTGAATCTCTCTCGAGCTGAAAAGGAG 2149
DB 121 ATTTTCATCATCGAGAGATCCACCGGGAGCTAAAGTGAATCTCTCTCGAGCTGAAAAGGAG 180
QY 2150 CATCTGGCTCAACTCGAATCATCACAGATTCGGAATTCAGATAGCAAACTTAGTCAGC 2209
DB 181 CATCTGGCTCAACTCGAATCATCACAGATTCGGAATTCAGATAGCAAACTTAGTCAGC 240
```

QY	2210	TGAAGACGACGAGTGGCGAGCGCTGCACATGAAGCAATAAATTTATTAGGAGGCA	2269	source	1..870
Db	241	TGAAGACGACGAGTGGCGAGCGCTGCACATGAAGCAATAAATTTATTAGGAGGCA	300		/organism="Homo sapiens"
QY	2270	AGAGGTACTGTTAGTAACTCTCAAGGAGAAATTTACGGTTGAGCACCAAAAAAGGAG	2329		/mol_type="mRNA"
Db	301	AGAGGTACTGTTAGTAACTCTCAAGGAGAAATTTACGGTTGAGCACCAAAAAAGGAG	360		/db_xref="taxon:9606"
QY	2330	TGATCATGAAGAAATATCAACAATTTATTAATGGTCAAGAGGAGTAGTCCGG	2389		/clone="IMAGE:6188752"
Db	361	TGATCATGAAGAAATATCAACAATTTATTAATGGTCAAGAGGAGTAGTCCGG	420		/sex="male"
QY	2390	TCCTACCACAAATCAATCTCCACAAATTCATTTGCTTTGAATGAACACACAGACAGAGAAG	2449		/tissue_type="sympathetic trunk"
Db	421	TCCTACCACAAATCAATCTCCACAAATTCATTTGCTTTGAATGAACACACAGACAGAGAAG	480		/lab_host="DH10B"
QY	2450	ACCATGATAAGAGAAGGCATCTTGACATAAAGTTCGTCTGACTCCAGCAGGAGATTCA	2509		/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1: Directionally cloned using the following adaptors: 5'-TGACACCTGCTAGTACGAGCGCGCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."
Db	481	ACCATGATAAGAGAAGGCATCTTGACATAAAGTTCGTCTGACTCCAGCAGGAGATTCA	540		
QY	2510	AATGGACGGTCTGTCATGGTCCCAAGTCTTACCATACTCTGAGACTGACTA	2569		
Db	541	AATGGACGGTCTGTCATGGTCCCAAGTCTTACCATACTCTGAGACTGACTA	600		
QY	2570	TCACCCAATTAGAAAAACAATCCCTTCATCCCTTTTTCATCCCACTGGGCATCATA	2629		
Db	601	TCACCCAATTAGAAAAACAATCCCTTCATCCCTTTTTCATCCCACTGGGCATCATA	660		
QY	2630	GGGCAAAATGGATCAAGGAGTTCAGATGTAGCAAAACCCAGAGAAATTTGCTTGGCTT	2689		
Db	661	GGGCAAAATGGATCAAGGAGTTCAGATGTAGCAAAACCCAGAGAAATTTGCTTGGCTT	720		
QY	2690	TAGCCATTTTGGAGTGTGCAGTTAAACCCAGTTCGTGATGCTACCAATATGCGGAGAATTT	2749		
Db	721	TAGCCATTTTGGAGTGTGCAGTTAAACCCAGTTCGTGATGCTACCAATATGCGGAGAATTT	780		
QY	2750	TAGGACATACCGGTTTACACCGGATGACATCAATT-GAAAGAGAAGAAAGGAGAAGTC	2808		
Db	781	TTAGACATACCGGTTTACACCGGATGACATCAATTGGAAAGAGAAGAAAGGAGNAAGT	840		
QY	2809	AAAAAAAAGAGAGAAACAGG	2830		
Db	841	CAAAAAAANAANAANAAG	862		
RESULT 9					
LOCUS	BQ718666	870 bp	mRNA	linear	EST 16-JUL-2002
DEFINITION	AGENCOURT_8234641	Lupski_sympathetic_trunk	Homo sapiens	cDNA clone	
ACCESSION	BQ718666				
VERSION	1	(bases 1 to 870)			
KEYWORDS	EST				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	NTH-MGC http://mgc.nci.nih.gov/ .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapb-r@mail.nih.gov				
	Tissue Procurement: Dr. James R. Lupski				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	Plate: LLAM13584	row: n	column: 17		
	High quality sequence stop: 671.				
FEATURES	Location/Qualifiers				

QY	7978	AAACAGGACGCAAAAAACGGAAGCGTGAAGAGAGTGTGGAGCAGAAACGTAAGCAAGCAG	8037		
Db	1	AAACAGGACGCAAAAAACGGAAGCGTGAAGAGAGTGTGGAGCAGAAACGTAAGCAAGCAG	60		
QY	8038	AATGCCACTAAGCTGTACGCTCTGCTTCAAGCACAAGAGCAGCTCAGAGCCGAGATC	8097		
Db	61	AATGCCACTAAGCTGTACGCTCTGCTTCAAGCACAAGAGCAGCTCAGAGCCGAGATC	120		
QY	8098	CTGAAGAAGAGAGCAGCTCTGGACAGAGATCTGCAAAATGAAGTGCAGGAAGAGTGAAG	8157		
Db	121	CTGAAGAAGAGAGCAGCTCTGGACAGAGATCTGCAAAATGAAGTGCAGGAAGAGTGAAG	180		
QY	8158	AGAGACCTGAAATTAAGAAAAAGAACCTGTAGTACGTTGGCTCAGGCCACAGAGTA	8217		
Db	181	AGAGACCTGAAATTAAGAAAAAGAACCTGTAGTACGTTGGCTCAGGCCACAGAGTA	240		
QY	8218	GCTGACCCCTGCCCTCCAGTGCACCCAGTTCCTCCAGCCCTCCAGCCCTCCACCTCA	8277		
Db	241	GCTGACCCCTGCCCTCCAGTGCACCCAGTTCCTCCAGCCCTCCAGCCCTCCACCTCA	300		
QY	8278	CCTCCCTCCCTCCAGTGTGTGCAACACACAGGCCCTTCTGTCCAGCCCTCCAGCTGT	8337		
Db	301	CCTCCCTCCCTCCAGTGTGTGCAACACACAGGCCCTTCTGTCCAGCCCTCCAGCTGT	360		
QY	8338	GCTTCCCAAGAGGAGCGGGAAGAGAAAGACTCCAGCTCAAGTCCAGAGAAAAG	8397		
Db	361	GCTTCCCAAGAGGAGCGGGAAGAGAAAGACTCCAGCTCAAGTCCAGAGAAAAG	420		
QY	8398	AAAATGATCTCTACTACTCTCAAGGAACTTAAGAGGACACAAAGCTTTACTGTATCTGT	8457		
Db	421	AAAATGATCTCTACTACTCTCAAGGAACTTAAGAGGACACAAAGCTTTACTGTATCTGT	480		
QY	8458	AAAACGCCCTTATGATGAATCTAAATTTATTTATTTGGCTGTGATCGGTGTGATCGTAC	8517		
Db	481	AAAACGCCCTTATGATGAATCTAAATTTATTTATTTGGCTGTGATCGGTGTGATCGTAC	540		
QY	8518	CATGGGGCTGCGTGTGGCATCTTGGCAAGTGAGGAGAGCTCAATGATGATGTCTGT	8577		
Db	541	CATGGGGCTGCGTGTGGCATCTTGGCAAGTGAGGAGAGCTCAATGATGATGTCTGT	600		
QY	8578	CCACAGTGGCCAGTCAACAGAGAGTGCATGACAGTGTCTCACGCCACTAACACAGAGAAGAT	8637		
Db	601	CCACAGTGGCCAGTCAACAGAGAGTGCATGACAGTGTCTCACGCCACTAACACAGAGAAGAT	660		
QY	8638	TATGA-GGGGTTGAAGAGGGTGTCTCGTCTCTTACAGGCCCATTAAGATGCCCTGGCTTT	8696		

```

Db      661  TATGAGGGGTTGAAGAGGTCTCCGTTCTTACAGCCCAATAGATGGCGCTTGCCTTT 720
QY      8697  CCTTGAACAGTAGACCCCTAATGATGCACCAATATTATGGTGTATTAAAGAACCTAT 8756
Db      721  CCTTGAACAGTAGACCCCTAATGATGCACCAATATTATGGGCTATTAAAGAACCTAT 780
QY      8757  GGACCTTCCACCATGGGAAGAGAGTACAAGACGATATTATGAAGAGCTG--ACGGAA 8814
Db      781  GGACCTTCCACCATGGGAAGAGAGTACAAGACGATATTATGAAGAGCTGGACGGAA 840
QY      8815  TTTGTGGCAGATATGACCAAAATTTT 8841
Db      841  TTGGGGGAGATATGACCAAAATTTT 867

RESULT 10
BU507358
LOCUS   BU507358      842 bp      mRNA      linear      EST 12-SEP-2002
DEFINITION AGENCOURT_10309897 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6501281
5', mRNA sequence.
ACCESSION BU507358
VERSION   BU507358.1 GI:22813591
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14055 row: h column: 18
High quality sequence stop: 751.
FEATURES             source
    source
        1..842
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6501281"
            /tissue_type="leiomyosarcoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_71"
            /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 2.1 kb."
BASE COUNT  257 a 246 c 171 g 167 t 1 others
ORIGIN
Query Match      8.1%; Score 800.2; DB 13; Length 842;
Best Local Similarity 99.4%; Pred. No. 2e-100;
Matches 813; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY      6129  TGGATCTCCAGTACAGTAACATTCACAAACAAAGAACTTTCATCAAACTTTTGCTAC 6188
Db      1  TGGATCTCCAGTACAGTAACATTCACAAACAAAGAACTTTCATCAAACTTTTGCTAC 60
QY      6189  ATGGGTTAAGCAAGCCAGTCAATTCAGGCTTGTTCAGTACAGCAAGTCTCTGG 6248
Db      61  ATGGGTTAAGCAAGCCAGTCAATTCAGGCTTGTTCAGTACAGCAAGTCTCTGG 120
QY      6249  TATCATTCATCAAGTACAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAG 6308
Db      121  TATCATTCATCAAGTACAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAG 180

```

FEATURES

```

QY      6309  AGCAACAGTCAAAATTAGGCCCAATACCTCAGGCTCTGGAGGAACACAAAGCAATTACA 6368
Db      181  AGCAACAGTCAAAATTAGGCCCAATACCTCAGGCTCTGGAGGAACACAAAGCAATTACA 240
QY      6369  AGTAATACAGAGCCCTCAGATTTCGCCCTGGTATGACCGTGATTAGAACACCACTCCA 6428
Db      241  AGTAATACAGAGCCCTCAGATTTCGCCCTGGTATGACCGTGATTAGAACACCACTCCA 300
QY      6429  GTCAACACTAGGAAAGGCAATTTATTTCGAACACCTGTGATGGTACGCCAGGTCTCTCA 6488
Db      301  GTCAACACTAGGAAAGGCAATTTATTTCGAACACCTGTGATGGTACGCCAGGTCTCTCA 360
QY      6489  GCAAGTGATGACTCAAAATCATCAGGGGGCAGCCTGTCTCCACTGAGTCTCGCCCTAA 6548
Db      361  GCAAGTGATGACTCAAAATCATCAGGGGGCAGCCTGTCTCCACTGAGTCTCGCCCTAA 420
QY      6549  CACGGTTTCCCTCAACACACCTTGGCAGAAAAGCTTAACCTTCAGCAACGTCACATTCA 6608
Db      421  CACGGTTTCCCTCAACACACCTTGGCAGAAAAGCTTAACCTTCAGCAACGTCACATTCA 480
QY      6609  ACAGTCTTCAGCCTCACAACACCCCTCGCCCCCAACAAAGGACAAAGTGAAGCTCACCAT 6668
Db      481  ACAGTCTTCAGCCTCACAACACCCCTCGCCCCCAACAAAGGACAAAGTGAAGCTCACCAT 540
QY      6669  TCAACTTACTCAGTTAAACACAGGGCCACGGTGGCAATCAAGTTTGACAGTAGTAAATCA 6728
Db      541  TCAACTTACTCAGTTAAACACAGGGCCACGGTGGCAATCAAGTTTGACAGTAGTAAATCA 600
QY      6729  AGGACAAGGTCAAACTACTGGACAGTTGCAGTTGATACCTCAAGGGGTGACTGTACTGCC 6788
Db      601  AGGACAAGGTCAAACTACTGGACAGTTGCAGTTGATACCTCAAGGGGTGACTGTACTGCC 660
QY      6789  AGGCCAGGCCAGCAGCTAAATGCAAGCTGCAATGCCAATGTACTGTTTCACGGATTCTCT 6848
Db      661  AGGCCAGGCCAGCAGCTAAATGCAAGCTGCAATGCCAATGTACTGTTTCACGGATTCTCT 720
QY      6849  CTTTACCCATTGGCAACAACAGCCACACAGCCAGCCACCAACACACCTGTTTTCAC 6908
Db      721  CTTTACCCATTGGCAACAACAGCCACACAGCCAGCCACCAACACACCTGTTTTCAC 780
QY      6909  GACAGCAGCAGGTACA-GGTGAACAAGGCAGAGTAAA 6945
Db      781  GACAGCAGCAGGTACAGGGTGAACAAGGCAGAGTAAA 818

RESULT 11
BU176205
LOCUS   BU176205      883 bp      mRNA      linear      EST 04-SEP-2002
DEFINITION AGENCOURT_7827724 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6139068
5', mRNA sequence.
ACCESSION BU176205
VERSION   BU176205.1 GI:22690189
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13455 row: h column: 13
High quality sequence stop: 662.
Location/Qualifiers

```


TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 1200)
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Inotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saich,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://phantom.gsc.riken.go.jp/ Location/Qualifiers 1. .1200 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:C330043E15" /db_xref="taxon:10090" /clone="C330043E15" /cell_type="ES cells" /clone_lib="RIKEN full-length enriched mouse cDNA library" 3/. .336 /note="unnamed protein product; putative unclassifiable" /codon_start=1 /protein_id="BAC38653.1" /db_xref="GI:26350027" /translation="MAWPELEVPDNDADYYGVKEPMDLATMEERIOKRYYEKLTE FVDMTKIFDNCRYNPRDTPFYQAEVLESFVOKLGFKASRSHNNKLQSTAP" 1176. .1181 polyA_signal /note="putative" 1200 polyA_site /note="putative" 371 a 241 c 243 g 345 t BASE COUNT ORIGIN Query Match 8.0%; Score 788.2; DB 11; Length 1200; Best Local Similarity 84.9%; Pred. No. 8.2e-99; Matches 1031; Conservative 0; Mismatches 153; Indels 31; Gaps 12; QY 8647 TTGAAGAGGTCCTCGCTTCCCTACAGGCCCAATAGATGCGCTTTCCTTGAACCA 8706 DB 1 TTGAAGAGGTCCTCGCTTCCCTACAGGCCCAAGATGCGCTTTCCTTGAACCG 60 QY 8707 GTAGACCCCAATGATGACCCAGATATTATGTTGTTATTAAAGAACCTATGACCTTGGC 8766 DB 61 GTAGACCCCAATGATGACCCAGATATTACGGTGTATTAAAGAGCAATGGACCTTGGC 120 QY 8767 ACCATGGAAGAAGAGTACAAAGCATATTATGAAGCTGACGCAATTTGTGGCAGAT 8826 DB 121 ACCATGGAAGAAGAGTACAAAGCGTATTATGAAGAAGCTGACAGAGTCTGGCGAGAT 180
QY	8827 ATGACCAAAATTTTGTATAACTGTGCTTACTACAATCAAGTGACGTCCCATTTTACCAG 8886
DB	181 ATGACCAAAATTTTGTATAACTGTGCTTACTACAATCAAGTGACGTCCCATTTTACCAG 240
QY	8887 TGTGCAGAAGTCTCGAATCATCTTTGTACAGAAATTTGAAGGCTTCAAGGTAGCAGG 8946
DB	241 TGTGCAGAAGTCTCGAATCATCTTTGTACAGAAATTTGAAGGATTTCAAGGCCAGCAGG 300
QY	8947 TCTCATAAACAACAACTGCAGCTTCTTAAAGTCTTCAAGCTGTTTAAACCTAACATA 9006
DB	301 TCTCATAAACAACAACTGCAGCTTCTTAAAGTCTTCAAGCTGTTTAAACCTAACATA 360
QY	9007 AAACACGCAAGAATCTGTTGTGAACATTTTAAATAAGAGCCGAGATGTTTATAG 9066
DB	361 CTAGACACGACAGTCTGGCGCTCTGAACTA-TTTAAACTAAAGCCAGATATTTTCAG 419
QY	9067 TCAGGCTATCCTTGACAAGACTTGACCTAACTTCTGTTTATTGGTCATACAACTCCAAT 9126
DB	420 TCAGGCTATCCTTGACAAGACTTGACCTAACTTCTGTTTATTGGTCATACAACTCCAAT 474
QY	9127 TATATTTCTTGGCCAAATTTTGTCCAGCGGCAAGAAAAAGCAAGTCAACGACACCATTA 9186
DB	475 TGTATTTCTTGGCCAAATTTTGTCCAGCGGCAAGAAAAAGCAAGTCAACGACACCATTA 534
QY	9187 TCTTGTCAAGATCAGATGTTTGTGTCAGAGGAGGAGAAACTTTTGTATTATGA 9246
DB	535 TCTTGTGAGAGCAAAATGGCTTTACTATTGTGCGAGAGCAGGAAACTTTTGTATTATGA 594
QY	9247 AAAAAAAGAAAAAGAAAG-----CAAGAAAAAAGATACTATGGGTCAAGT 9294
DB	595 AAAAAAAGAAAAAAGAAAAAGCAAGAAAAAAGAAAGATACTATGGGTCAAGT 654
QY	9295 GTAACCTCCATGGAAATGCCAGCTCTCTTCAAGTGAAGAGTGGTTAGAGTCTCACA 9354
DB	655 GTAACCT-CGGGGGGATGCCACATCTCTTCAAGTGAAGAGTGGTTAGAGTCTCACA 712
QY	9355 GAAAACCTTTGACTGTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 9412
DB	713 -AAACTTTTGTGACTGTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 771
QY	9413 CTCATTTGTCAGCTAGTATTATTTCTTATA-AAATCCAGCCCGGTGTACATATAATCA- 9470
DB	772 CTCATTTGTCAGCTAGTATTATTTTCTTATACAAATCCAGCCGTGCGTGTATCCAC 831
QY	9471 TCTGTATCTATCATGATTCCTGTAGGTAAAGTACAAGACGACCTCTAGATGCTCTTTC 9530
DB	832 CCGTATCTATCATGATTCCTGTAGGT-GAAGTACAAGCGGACCTCTAGTGTCTTTTC 890
QY	9531 TTTCTATGAAGAGGAGTGTATGTACACATGTGCACACACACACAACTGGGAATCAACA 9590
DB	891 TTTCTATGAAGAGGAGTGTATGTACACATGTG---CACACACAACCTGGGAATCAACA 946
QY	9591 TGAAGTTTATTGTTTCATGTTAGATTAAATTAAGCTTGCATAAAGTTGGCTAAGTGGTC 9650
DB	947 TGAAGTTTATTGTTTCATGTTAGATTAAAGCTTGCATAAAGTTGGCTAAGTGGTC 1006
QY	9651 CTGCGGTACAGACTCTGTTGCTTGAATATAACAGTACAAATTTGTCAAATTTACTCTGCAC 9710
DB	1007 CTGCGGTACAGACTCTGTTGCTTGAATATAACAGTACAAATTTGTCAAATTTACTCTGCAC 1066
QY	9711 CAGGCTAAAGTGAAGTAAATCTATTGTAAGGTATCTTGTGTTGTAACATTTGTGAGATTC 9770
DB	1067 CAGGCTAAA-TGAGCAAAATCTATTGTAAGGTATCTTGTGTTGTAACATTTGTGAGATTC 1125
QY	9771 TAATTTTTTCTTTTGTATTAATTTCAACTATGATGATATATGAACAAATTAATATGA 9830
DB	1126 AATTTTTTCTTTTGTATTAATTTCAACTATGATGATGATATATGAACAAATTAATATGA 1185
QY	9831 GATAATTTTCTCTCC 9845
DB	1186 GATCATTTTCTCTC 1200

```

RESULT 13
BQ717925          864 bp      mRNA      linear      EST 16-JUL-2002
LOCUS             AGENCOURT_8241359 Lupski_sympathetic_trunk Homo sapiens cDNA clone
DEFINITION        IMAGE:6187125 5', mRNA sequence.

ACCESSION        BQ717925
VERSION          BQ717925.1 GI:21856822
KEYWORDS         EST.
SOURCE           Homo sapiens (human)
ORGANISM         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE        NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS          National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE            Unpublished
JOURNAL          Contact: Robert Strausberg, Ph.D.
COMMENT         Email: cgapbs-ref@mail.nih.gov
                  Tissue Procurement: ATCC
                  cDNA Library Preparation: Life Technologies, Inc.
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Agencourt Bioscience Corporation
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLAM13580 row: j column: 22
                  High quality sequence stop: 628.
FEATURES         Location/Qualifiers
                  1..864
                    source          /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:6187125"
                    /sex="male"
                    /tissue_type="sympathetic trunk"
                    /dev_stage="adult, 16 yr"
                    /lab_host="DH10B"
                    /clone_lib="Lupski_sympathetic_trunk"
                    /note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
                  NotI; Site_2: SalI; cDNA made by oligo-dT priming.
                  Directionally cloned using the following adaptors:
                  5'-TCGACCACAGGTCGG-3' and
                  5'-GACTAGTCTAGATCGGAGCGGCCCT(15)-3'. Size selected >
                  1 kb for average insert length 1.9 kb. This is a primary
                  library, non-amplified. Library constructed by Life
                  Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
                  College of Medicine); available through Life
                  Technologies."
BASE COUNT       254 a 298 c 154 g 157 t 1 others
ORIGIN
Query Match      7.9%; Score 781.2; DB 13; Length 864;
Best Local Similarity 98.0%; Pred. No. 8.3e-98;
Matches 823; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

QY 6546 TAACACGGTTTCCTCAACACCTGGGCGAGAAAGCTTAACCTCAGACAGCTCCACTTCAA 6605
Db      |||||||
QY 6606 TATACAGTCTCAGCTCACAACCCCTCGCCCCCAACAGCAAGTGAAGTCAACCAT 6665
Db      |||||||
QY 6666 GGCTCAACTTACTCAGTTTAACACAGGGCGACGGTGGCAATCAAGTTTGACAGTAGTAAT 6725
Db      |||||||
QY 6725 TATACAGTCTCAGCTCACAACCCCTCGCCCCCAACAGCAAGTGAAGTCAACCAT 120
Db      |||||||
QY 6726 TCAAGGACAAGTCAAACTACTGGACAGTTGCAGTTGCATCTCAAGGGGTGACTGTACT 6785
Db      |||||||
QY 6785 TCAAGGACAAGTCAAACTACTGGACAGTTGCAGTTGCATCTCAAGGGGTGACTGTACT 240
Db      |||||||
QY 6786 CCCAGGCCAGGCCAGCAGCTAAATGCAAGCTGCAATGCCAATGGTACTGTTTCAGCGGATT 6845
Db      |||||||

```

```

241 CCACAGCCAGCCAGCAGCTAATGCAAGCTGCAATGCCAATGGTACTGTTTCAGCGATT 300
6846 CTTCTTTACCCCATTTGGCAACACAGCCACAGCCAGCCAGCCAGCCAGCCAGCCAGCTTTTC 6905
301 CTTCTTTACCCCATTTGGCAACACAGCCACAGCCAGCCAGCCAGCCAGCCAGCTTTTC 360
6906 CACGACAGCAGCAGGTACAGGTGAACAAAGGAGAGATAAAGTGTCAACCCAGATGAGGT 6965
361 CACGACAGCAGCAGGTACAGGTGAACAAAGGAGAGATAAAGTGTCAACCCAGATGAGGT 420
6966 ACATCAAGCAAAACCCCTGCCACAGCTCAGTCAAGTGTGGTCCAGCAAAAGCCCA 7025
421 ACATCAAGCAAAACCCCTGCCACAGCTCAGTCAAGTGTGGTCCAGCAAAAGCCCA 480
7026 GCCACAGACTGCTCAGCCTTTCAGTCTGGCCGCCAGCCCAAAAGCCAGTCCAGCAGC 7085
481 GCCACAGACTGCTCAGCCTTTCAGTCTGGCCGCCAGCCCAAAAGCCAGTCCAGCAGC 540
7086 TCAGCTTGAAGTTTCAGACTCAGCTGAGTTTCAGACCCCAAAAGCTGTTTCATCCCATGT 7145
541 TCAGCTTGAAGTTTCAGACTCAGCTGAGTTTCAGACCCCAAAAGCTGTTTCATCCCATGT 600
7146 CCCTTCTGAAGCACAACCCAGCCAGCAGTTCATCCAAAGCCCAAAAGTTGCAGCAGATC 7205
601 CCCTTCTGAAGCACAACCCAGCCAGCAGTTCATCCAAAGCCCAAAAGTTGCAGCAGATC 660
7206 TCAGCTTCAAAAGTAATGTCCAAAGGAGAGTCTCTGTGTGTCCTCAAGTCCATCAGAC 7265
661 TCAGCTTCAAAAGTAATGTCCAAAGGAGAGTCTCTGTGTGTCCTCAAGTCCATCAGAC 7322
7266 TCAGCTTCAAAAGTAATGTCCAAAGGAGAGTCTCTGTGTGTCCTCAAGTCCATCAGAC 7381
721 TCAGCTTCAAAAGTAATGTCCAAAGGAGAGTCTCTGTGTGTCCTCAAGTCCATCAGAC 780
7323 GACTTACAACCTCAC- AACCGATTCCAAATTCACACACATACATCTCTTCAGATACCTTCCC 7381
781 GACTTACAACCTCAC- AACCGATTCCAAATTCACACACATACATCTCTTCAGATACCTTCCC 840

BM451623          1050 bp      mRNA      linear      EST 05-FEB-2002
AGENCOURT_6394955 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493133
5', mRNA sequence.
ACCESSION        BM451623
VERSION          BM451623.1 GI:18500663
KEYWORDS         EST.
SOURCE           Homo sapiens (human)
ORGANISM         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE        NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS          National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE            Unpublished
JOURNAL          Contact: Robert Strausberg, Ph.D.
COMMENT         Email: cgapbs-ref@mail.nih.gov
                  Tissue Procurement: ATCC
                  cDNA Library Preparation: Life Technologies, Inc.
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Agencourt Bioscience Corporation
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLAM12116 row: f column: 14
                  High quality sequence stop: 641.
FEATURES         Location/Qualifiers
                  1..1050
                    source          /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5493133"
                    /tissue_type="retinoblastoma"

```



```
/lab_host="DH10B (phage-resistant)"
/clone_lib="Nih_Mc67"
/Note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT      341 a   189 c   270 g   248 t   2 others
ORIGIN

Query Match      7.9%; Score 778; DB 12; Length 1050;
Best Local Similarity 92.1%; Pred. No. 2.2e-97;
Matches 877; Conservative 0; Mismatches 62; Indels 13; Gaps 5;

QY 648 CTTTTCGAGCTCTGGTGAGCCAGAGCAGTGCACACTCATGCGCAGAGATGCATGTTGT 707
Db 13 CTTTTCGAGCTCTGGTGAGCCAGAGCAGTGCACACTCATGCGCAGAGATGCATGTTGT 72
QY 708 GCTTTTGAAGCAGTCTCGGTGAAGACACACTTCCAACTACTACCTTTGGACCTGCTGA 767
Db 73 GCTTTTGAAGCAGTCTCGGTGAAGACACACTTCCAACTACTACCTTTGGACCTGCTGA 132
QY 768 TCTGAAAGATAGCGTTAATTCACACTGTATTTTCATAGATGGGATGACGTGGCCAGAGGT 827
Db 133 TCTGAAAGATAGCGTTAATTCACACTGTATTTTCATAGATGGGATGACGTGGCCAGAGGT 192
QY 828 GCTGCGGGTGACTGTGAGAGTGATTAAGGAGTACCATCAGTTCCTTCCATTACCAAGAGGC 887
Db 193 GCTGCGGGTGACTGTGAGAGTGATTAAGGAGTACCATCAGTTCCTTCCATTACCAAGAGGC 252
QY 888 AGAGGACTACCATATGGACAGTACAGAACAGATCAAAAGTCTACAGTTCTAGTCCGA 947
Db 253 AGAGGACTACCATATGGACAGTACAGAACAGATCAAAAGTCTACAGTTCTAGTCCGA 312
QY 948 TCAGTTTCTTACAACAATATTCCTCGAGAGGAATGATGCTCGAAGGGGTGATACAGTA 1007
Db 313 TCAGTTTCTTACAACAATATTCCTCGAGAGGAATGATGCTCGAAGGGGTGATACAGTA 372
QY 1008 TGATGACCAATGTAGGGTTTGTCACAACTTGGGGATTTGCTTGTGTGAGACATGTTTC 1067
Db 373 TGATGACCAATGTAGGGTTTGTCACAACTTGGGGATTTGCTTGTGTGAGACATGTTTC 432
QY 1068 AGCAGTATACCAATTTGGAATGTGTGAAGCCACTCTTGAGGAGGTGCCAGAGGACAGTG 1127
Db 433 AGCAGTATACCAATTTGGAATGTGTGAAGCCACTCTTGAGGAGGTGCCAGAGGACAGTG 492
QY 1128 GCAGTGTGAAGTCTGTGTAGCACACAAGGTGCTGTGTGACTGACTGTGTGCTGAAT 1187
Db 493 GCAGTGTGAAGTCTGTGTAGCACACAAGGTGCTGTGTGACTGACTGTGTGCTGAAT 552
QY 1188 CCAAAAAATAAACCATATATTTCGACATGAACCTATTGGATATGATAGAAGTCGGAGAA 1247
Db 553 CCAAAAAATAAACCATATATTTCGACATGAACCTATTGGATATGATAGAAGTCGGAGAA 612
QY 1248 ATACTGTTCTTGAACCGAGACTCATATAGAGAGAGATACAGAAAATGAAAATGAAA 1307
Db 613 ATACTGTTCTTGAACCGAGACTCATATAGAGAGAGATACAGAAAATGAAAATGAAA 672
QY 1308 GAAAATTTGGTATTACAGCACAAAAGGTCCAACTTGAGAAATTAATGTACTGTCTAGACAA 1367
Db 673 GAAAATTTGGTATTACAGCACAAAAGGTCCAACTTGAGAAATTAATGTACTGTCTAGACAA 732
QY 1368 AGATTATT--GGGAAGCAGAACTCTGCAAAATTTCTAGAGAAATGCGTGAAGAAA----TC 1422
Db 733 AGATTATTGGGAAGCAGAACTCTGCAAAATTTCTAGAGAAATGCGTGAAGAAATCCAC 792
QY 1423 CACCGACACATGGACATACTAGAGACCTGACCAATAAG--CTCGGGCGAGTACAAA 1479
Db 793 CCGACACATGGNGACATAACTGAAGACCTGACCCCAATAAGGCTCGGGGGCGAGTACAAA 852
QY 1480 TCC-TTTCGCGGCGAGC---TAATGAAGAAATTTTGGAAATTCATTAAGAGCCAAAAGG 1534
Db 853 TCCTTTTCGCGGCGAGCCTAATGGGAAGAAATTTTGGAAATCCCTTAAGAGACCAANAAG 912
```

```
QY 1535 GAGACATTGATATGTTTAAAGCCCGAGAGAAACAGAAAAAGACACAGAAATGA 1586
Db 913 GGGAGACATTTGTAATTAATGTTTAAAGCCCGAGAGAAACCGGAAANA 964

RESULT 15
LOCUS      AU124130
DEFINITION AU124130 NT2RM2 Homo sapiens cDNA clone NT2RM2001716 5', mRNA
sequence.
ACCESSION AU124130
VERSION    AU124130.1 GI:10948846
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 841)
AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
TITLE      HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
Y., Sugano,S., Isogai,T.)
JOURNAL    Unpublished
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES
     source
       1..841
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="NT2RM2001716"
         /cell_type="teratocarcinoma"
         /cell_line="NT2"
         /clone_lib="NT2RM2"
         /notes="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
         precursor cells"
BASE COUNT  276 a   204 c   189 g   169 t   3 others
ORIGIN

Query Match      7.8%; Score 771.4; DB 9; Length 841;
Best Local Similarity 98.2%; Pred. No. 1.9e-96;
Matches 800; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 5682 AATTAGAGGAGAGATGTTGGTCTTATGGCATTCGATTGTAATTTGTATCAGGAAAT 5741
Db 2 AATTAGAGGAGAGATGTTGGTCTTATGGCATTCGATTGTAATTTGTATCAGGAAAT 61
QY 5742 CATTTCTCCCATTTGGAGTTCACAGAAACACCAAAAGAACCCCTACACCTCAGAGGAAGG 5801
Db 62 CATTTCTCCCATTTGGAGTTCACAGAAACACCAAAAGAACCCCTACACCTCAGAGGAAGG 121
QY 5802 CCTTCGATCAAGTGCACCTCGGCCAAAGAGACCAGAAACGCCCAAGCAAACTGCCCTGT 5861
Db 122 CCTTCGATCAAGTGCACCTCGGCCAAAGAGACCAGAAACGCCCAAGCAAACTGCCCTGT 181
QY 5862 TATTATTGAACCTCGGTAGCAGAAAGAACTGGAATTTGTGGAGATCAGGCAATTCG 5921
Db 182 TATTATTGAACCTCGGTAGCAGAAAGAACTGGAATTTGTGGAGATCAGGCAATTCG 241
QY 5922 TGAGAGAGTGGAGAAAGAAAGCCACAGCAGTTGAGCACACAGGCTTAAGAAACGACTGGA 5981
Db 242 TGAGAGAGTGGAGAAAGAAAGCCACAGCAGTTGAGCACACAGGCTTAAGAAACGACTGGA 301
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:12:14 ; Search time 404 Seconds
(without alignments)
10777.833 Million cell updates/sec

Title: US-09-698-295-2
Perfect score: 9865
Sequence: 1 ggcgcggctgagtggtgccc.....acaaaaaaaaaaaaaaaaa 9865

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	814.6	8.3	1032	4	US-09-257-179-21
2	101.8	1.0	1926	4	US-09-249-585A-4
3	101.8	1.0	1931	2	US-09-130-114-2
4	93.2	0.9	7218	1	US-08-232-463-14
5	89	0.9	390	3	US-09-197-649-7
6	85.2	0.9	3489	2	US-08-728-323A-1
7	85.2	0.9	3489	4	US-09-298-568-1
8	85.2	0.9	3489	4	US-09-410-399-1
9	85.2	0.9	32207	2	US-08-770-379-20
10	85.2	0.9	32207	3	US-08-757-669A-20
11	85.2	0.9	32207	4	US-09-230-371A-20
12	81.4	0.8	124884	4	US-09-661-596A-76
13	79	0.8	51259	3	US-08-781-891-209
14	79	0.8	51259	4	US-09-618-166-209
15	77.6	0.8	1926	4	US-09-249-585A-2
16	77.6	0.8	1926	4	US-09-410-399-3
17	77.6	0.8	2580	3	US-09-050-863-2
18	77.6	0.8	2580	4	US-09-359-081-2
19	77.6	0.8	5452	2	US-09-130-114-1
20	77.6	0.8	8705	4	US-09-647-344A-14
21	77.6	0.8	9600	3	US-08-910-647-1
22	77.6	0.8	9600	4	US-09-620-925-1
23	77.6	0.8	10596	1	US-07-884-811-15
24	77.6	0.8	10596	1	US-07-885-971-15
25	77.6	0.8	10596	1	US-08-087-783A-15
26	77.6	0.8	10596	1	US-08-194-088B-15
27	77.6	0.8	10596	2	US-08-194-087-15

28	77.6	0.8	10596	5	PCT-US93-04648-15	Sequence 15, Appl
c 29	77	0.8	16442	3	US-08-781-891-208	Sequence 208, App
c 30	77	0.8	16442	4	US-09-618-166-208	Sequence 208, App
31	68.6	0.7	4549	4	US-09-620-312D-266	Sequence 266, App
32	68.6	0.7	4942	4	US-09-620-312D-267	Sequence 267, App
33	66.6	0.7	2277	1	US-08-676-967-5	Sequence 5, Appl
34	66.6	0.7	2277	1	US-08-676-967-5	Sequence 5, Appl
35	66.6	0.7	2277	2	US-09-098-487-5	Sequence 5, Appl
c 36	66.6	0.7	2481	4	US-09-894-998A-35	Sequence 35, Appl
37	65.8	0.7	2301	1	US-08-306-691B-23	Sequence 23, Appl
38	65.8	0.7	2301	4	US-09-167-206-3	Sequence 3, Appl
39	65.8	0.7	2301	5	PCT-US93-06251-78	Sequence 78, Appl
40	65.6	0.7	2721	6	5215881-2	Sequence 1, Appl
41	65.6	0.7	8438	1	US-07-945-283-1	Sequence 8, Appl
42	61	0.6	150	2	US-07-829-461A-8	Sequence 21, Appl
c 43	60.6	0.6	5340	4	US-09-627-122-21	Sequence 1, Appl
44	60.6	0.6	15378	3	US-08-785-420-1	Sequence 1, Appl
45	60.4	0.6	436	4	US-09-397-787-257	Sequence 257, App

ALIGNMENTS

RESULT 1									
US-09-257-179-21									
; Sequence 21, Application US/09257179									
; Patent No. 6410709									
; GENERAL INFORMATION:									
; APPLICANT: Ruben et al.									
; TITLE OF INVENTION: 29 Human Secreted Proteins									
; FILE REFERENCE: P2015P1									
; CURRENT APPLICATION NUMBER: US/09/257,179									
; CURRENT FILING DATE: 1999-02-25									
; EARLIER APPLICATION NUMBER: PCT/US98/17709									
; EARLIER FILING DATE: 1998-08-27									
; EARLIER APPLICATION NUMBER: 60/056,270									
; EARLIER FILING DATE: 1997-08-29									
; EARLIER APPLICATION NUMBER: 60/056,271									
; EARLIER FILING DATE: 1997-08-29									
; EARLIER APPLICATION NUMBER: 60/056,247									
; EARLIER FILING DATE: 1997-08-29									
; EARLIER APPLICATION NUMBER: 60/056,073									
; EARLIER FILING DATE: 1997-08-29									
; NUMBER OF SEQ ID NOS: 128									
; SOFTWARE: PatentIn Ver. 2.0									
; SEQ ID NO 21									
; LENGTH: 1032									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-09-257-179-21									
Query Match 8.3%; Score 814.6; DB 4; Length 1032;									
Best Local Similarity 98.3%; Pred. No. 1.2e-167;									
Matches 823; Conservative 0; Mismatches 14; Indels 0; Gaps 0;									
QY	8444	TTTACTGTATCTGTAACGCCCTTATGATGAATCTAAATTTTATTTGCGTGTGATCGGT	8503						
Db	179	TGTTACTGTATCTGCAGAACACCTTATGATGAGTCACAAATTTTATTTGCGTGTGATCGGT	238						
QY	8504	GTCCAGATTTGGTACCATGGCGCTGGTGGGATCTTTGCAAGTGAGGCAGAGCTCATTTG	8563						
Db	239	GTCCAGATTTGGTACCATGGCGCTGGTGGGATCTTTGCAAGTGAGGCAGAGCTCATTTG	298						
QY	8564	ATCAGTATGTCTCTCCACAGTGCAGTCAACAGAGGATGCATGACAGTCTCTCAGCCAC	8623						
Db	299	ATCAGTATGTCTCTCCACAGTGCAGTCAACAGAGGATGCATGACAGTCTCTCAGCCAC	358						
QY	8624	TACAGAGAGGATTTATGAGGGGTGTAAGAGGGTGTCCGTTCTTACAGGCCATAAGA	8683						
Db	359	TACAGAGAGGATTTATGAGGGGTGTAAGAGGGTGTCCGTTCTTACAGGCCATAAGA	418						
QY	8684	TGGCCTGGGCTTTCTTGAACACAGTACACCCCTTAATGATGCACAGATTATTATGGTGA	8743						

Db 419 TGCGCTGGCCCTTCCTTGAACCACTAGACCCCTAATGATGCACACAGATATTATTGTTGTTA 478
Qy 8744 TTAAGGAACCTATGGACCTTGCCACCATGGAAGAAAGAGTACAAAGACGATATTATGAAA 8803
Db 479 TTAAGGAACCTATGGACCTTGCCACCATGGAAGAAAGAGTACAAAGACGATATTATGAAA 538
Qy 8804 AGCTGACGGAAATTTGTGGCAGATATGACCAAAATTTTTGATAACTGCTGTTACTCAATC 8863
Db 539 AGCTGACGGAAATTTGTGGCAGATATGACCAAAATTTTTGATAACTGCTGTTACTCAATC 598
Qy 8864 CAAGTACTCCCATTTTACCAGTGTGCAGAAAGTCTCGAATCATCTTTGTGACAAAT 8923
Db 599 CAAGTACTCCCATTTTACCAGTGTGCAGAAAGTCTCGAATCATCTTTGTGACAAAT 658
Qy 8924 TGAAGGCTTCAAGCTAGCAGTCTCATAAACAAACTGCACTGCTACAGCTTCTTAA 8983
Db 659 TGAAGGCTTCAAGCTAGCAGTCTCATAAACAAACTGCACTGCTACAGCTTCTTAA 718
Qy 8984 GTTCAGCGTGTAACTTAACATAAACAACAGCAAGAAATCTGTTGCTGAACATATTTAA 9043
Db 719 GTTCAGCGTGTAACTTAACATAAACAACAGCAAGAAATCTGTTGCTGAACATATTTAA 778
Qy 9044 ATTAAGAGCCAGATGTTTTTACTCAGGCTATCTGACAGACTTGACCTAATCTCGTT 9103
Db 779 ATTAAGAGCCAGATGTTTTTACTCAGGCTATCTGACAGACTTGACCTAATCTCGTT 838
Qy 9104 TTTATTGCTCATACAGTCCCAATATATTTCTGGCCAAATTTTGTCCAAACGCAAGAAAA 9163
Db 839 TTTATTGCTCATACAGTCCCAATATATTTCTGGCCAAATTTTGTCCAAACGCAAGAAAA 898
Qy 9164 AAGCAAGTCAAGCAGCACTATCTTGTCAAGATCATGATGTTTACTATTGTCGAGA 9223
Db 899 AAGCAAGTCAAGCAGCACTATCTTGTCAAGATCATGATGTTTACTATTGTCGAGA 958
Qy 9224 AGCAGAAACTTGTATTGAAAAAAGAAAAAGAAAGAAAGCAAGAAAAAGATA 9280
Db 959 AGCAGAAACTTGTATTGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1015

RESULT 2

US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match 1.0%; Score 101.8; DB 4; Length 1926;
Best Local Similarity 55.5%; Pred. No. 1.9e-12;
Matches 217; Conservative 0; Mismatches 172; Indels 2; Gaps 1;
Qy 6 GGCTGAGTGGCGCCCAAGACGGCGCTGAGTCCGCCAGGGGGGCGACGAGTACCCGAG 65
Db 760 GACGGGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGG 701
Qy 66 GAAGCCG 125
Db 700 GAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGAG 641

Qy 126 AGGAGCGCGGGGGCGGACGACGGG--GGCGGGGGGGGGCGGGCCACCTGTCCCGGACC 183
Db 640 GACGGGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGAG 581
Qy 184 ACCGCGCGCGGGGGCGGCTCAACAAAGTGTGTACGATGACACAGAGAGCGAGGCGGTG 243
Db 580 GACGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGAGGACGGG 521
Qy 244 GAGGAAGAGGAGGACATGCTCTCCGAGGAGGAGGAGGAGGACGGCGCCAGCCCGGAGGAG 303
Db 520 GAGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGAGGACGGGAGGAC 461
Qy 304 ACCAGGATCTGAGGACGACGAGGAGGATGAGATGGAAGAGGACGACGATGACTCCGAT 363
Db 460 GGGGAGGACGAGGACGAGGAGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAG 401
Qy 364 TATCCGAGGAGATGGAAGACGACGACGACG 394
Db 400 GAGGACGAGGACGGGGAGGAGGACGAGGACG 370

RESULT 3

US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5978807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903051
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match 1.0%; Score 101.8; DB 2; Length 1931;
Best Local Similarity 55.5%; Pred. No. 1.9e-12;
Matches 217; Conservative 0; Mismatches 172; Indels 2; Gaps 1;
Qy 6 GGCTGAGTGGCGCCCAAGACGCGGCTGAGTCCGCCAGGGGGGCGACGAGTACCGGAG 65
Db 760 GACGGGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGG 701
Qy 66 GAAGCCG 125
Db 700 GAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGAG 641
Qy 126 AGGAGCGGGGGCGGCGACGACGGG--GGCGGGGGGGCGGCGCGCGCGCGCGCGCGCGGACC 183
Db 640 GACGGGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGAGGACGAGGACGGGGAGG 581
Qy 184 ACCGCGCGCGGGGGCGGCTCAACAAAGTGTGTACGATGACCAAGAGACGAGGCGGTG 243
Db 580 GACGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGG 521
Qy 244 GAGGAAGAGGAGGACATGCTCTCCGAGGAGGAGGAGGAGGACGGCGCGGAGGAGGAGGAGGAG 303
Db 520 GAGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGAGGAGGAGGAC 461
Qy 304 ACCAGGATCTGAGGACGACGAGGAGGATGAGATGGAAGAGGACGACGATGACTCCGAT 363
Db 460 GGGGAGGAGGACGAGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGAGGAGGAGGAG 401
Qy 364 TATCCGAGGAGATGGAAGACGACGACGACG 394
Db 400 GAGGACGAGGACGGGGAGGAGGACGAGGACG 370

Db 783 AGGAGCAGGAGGGCCAGGAGGGCCAGGAGGAGGGCCAGGAGGGCCAGGAGCAGG 842
Qy 218 ACGATGACACGAGAGCGCGGTGGAGGAAGAGGAGGACATGGTCTCCGAGGAGGAGG 277
Db 843 AGGAGGGCCAGGAGGGCCAGGAGGAGGGCCAGGAGGGCCAGGAGGGCCAGGAGG 902
Qy 278 AGGAGGAGGAGCGGCGGAGCGGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 337
Db 903 AGGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 962
Qy 338 TGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 397
Db 963 GGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1022

Search completed: September 23, 2003, 09:19:26
Job time : 408 secs

Matches 339; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 7531 CTGCAACAACAGTCCAGGTTCTCTCTCAGATCCAGTCCAGGTTGTGGCTCAGATACAG 7590

Db 341 CTGCAACAACAGTCCAGGTTCTCTCTCAGATCCAGTCCAGGTTGTGGCTCAGATACAG 282

QY 7591 GCTCAGCAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATCCAAATTCAGCAA 7650

Db 281 GCTCAGCAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATCCAAATTCAGCAA 222

QY 7651 AGCAGTGTGTGCAGACTCACCAGATTCCAGATGTCAGATGTCAGGCGAGGCCAGTGTG 7710

Db 221 AGCAGTGTGTGCAGACTCACCAGATTCCAGATGTCAGATGTCAGGCGAGGCCAGTGTG 162

QY 7711 CAAGAGCAGTGTGCAAGGGTTCAGCAACTCAGGATCAGCAGCAAAAGAAAGCAAGCAA 7770

Db 161 CAAGAGCAGTGTGCAAGGGTTCAGCAACTCAGGATCAGCAGCAAAAGAAAGCAAGCAA 102

QY 7771 CAGATAGAAATTA-ACGTGAACACACACCCCTCCAGCTTCTAATCAAAAGTTGAATCATTTCA 7829

Db 101 CAGATAGAAATTAAGCGGTGAACACACCCCTCCAGCTTCTAATCAAAAG-TGAATCATTTCA 43

QY 7830 GAAACAGGTGGTGAAGCATAAATGCTGTAATAGAACATTT 7871

Db 42 GAAACAGGTGGTGAAGCATAAATGCTGTAATAGAACATTT 1

RESULT 13

US-09-796-692-8224

; Sequence 8224, Application US/09796692

; Publication No. US20020198362A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Mannion, Jane

; APPLICANT: Algate, Paul A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

; FILE REFERENCE: 2077.001200

; CURRENT APPLICATION NUMBER: US/09/796,692

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 60/222,903

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: 60/223,416

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: 60/223,378

; PRIOR FILING DATE: 2000-08-07

; NUMBER OF SEQ ID NOS: 9597

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 8224

; LENGTH: 545

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (419)

; OTHER INFORMATION: n=A,T,C or G

US-09-796-692-8224

Query Match 2.3%; Score 223.2; DB 10; Length 545;

Best Local Similarity 96.2%; Pred. No. 1.4e-40;

Matches 228; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 8444 TTTACTGTATCTGTTAAAGCGCTTATGATCAATCTAAATTTTATATTTGCTGTGATCGGT 8503

Db 309 TGTACTGTATCTGCGAACAACCTTATGATGAGTCACAATTTTATATTTGCTGTGATCGGT 368

QY 8504 GTCAGAAATTTGTTACCATGCGGCGCTGCGTTGGGATCTTGCAAAAGTGAGGCAGAGCTCATTTG 8563

Db 369 GTCAGAAATTTGTTACCATGCGGCGCTGCGTTGGGATCTTGCAAAAGTGAGGCAGAGCTCATTTG 428

QY 8564 ATGAGTATCTGTTCCACAGTCCAGTCAACAGAGATGCCATGACAGTGTTCACGCCAC 8623

Db 429 ATGAGTATCTGTTCCACAGTCCAGTCAACAGAGATGCCATGACAGTGTTCACGCCAC 488

QY 8624 TAACAGAGAAGGATTATGAGGGGTTGAAGAGGGTGTCTCGTTTCCCTTACAGGCCCATTA 8680

Db 489 TAACAGAGAAGGATTATGAGGGGTTGAAGAGGGTGTCTCGTTTCCCTTACAGGCCCATTA 545

RESULT 14

US-10-040-862-8224

; Sequence 8224, Application US/10040862

; Publication No. US20030078396A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; APPLICANT: Retter, Marc

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther

; FILE REFERENCE: Hematological Malignancies

; CURRENT APPLICATION NUMBER: US/10/040,862

; CURRENT FILING DATE: 2001-11-06

; PRIOR APPLICATION NUMBER: US 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: US 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: US 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: US 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: US 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: US 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: US 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: US 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: US 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: US 60/222,903

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: US 60/223,416

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: US 60/223,378

; PRIOR FILING DATE: 2000-08-07

; PRIOR APPLICATION NUMBER: US 09/796,692

; NUMBER OF SEQ ID NOS: 10467

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 8224

; LENGTH: 545

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: unsure

Search completed: September 24, 2003, 01:08:57
Job time : 1498 secs

```
; LOCATION: (419)
; OTHER INFORMATION: n=A,T,C or G
US-10-040-862-8224

Query Match      2.3%; Score 223.2; DB 14; Length 545;
Best Local Similarity 96.2%; Pred. No. 1.4e-40;
Matches 228; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 8444 TTACTGTATCTGTAAACGCCCTTATGATGAATCTAAATTTTATTTGGCTGTGATCGGT 8503
DB 309 TGTACTGTATCTGCAGACACCTTATGATGAGTCACAAATTTATTTGGCTGTGATCGGT 368

QY 8504 GTCAGAAATCGTACCATGGCGGTGGCTTGGCATCTTGCATCTTCAAAGTGAGGCAGAGCTCATTTG 8563
DB 369 GTCAGAAATCGTACCATGGCGGTGGCTTGGCATCTTGCATCTTCAAAGTGAGGCANAGCTCATTTG 428

QY 8564 ATGAGTATGCTGTCCACAGTGCAGTCACAGAGAGTGCATGACAGTGTCTCACGCCAC 8623
DB 429 ATGAGTATGCTGTCCACAGTGCAGTCACAGAGAGTGCATGACAGTGTCTCACGCCAC 488

QY 8624 TAACAGAGAAGGATTATGAGGGTTGAAGAGGGTGTCTCCCTTACAGGCCCATTA 8680
DB 489 TACAGAGAGGATTATGAGGGTTGAAGAGGGTGTCTCCCTTACAGGCCCATTA 545

RESULT 15
US-09-728-445-869
; Sequence 869, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; TITLE OF INVENTION: Animals
; FILE REFERENCE: LEX-0102--USA
; CURRENT APPLICATION NUMBER: US/09728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 869
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(405)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-445-869

Query Match      1.8%; Score 181.4; DB 10; Length 405;
Best Local Similarity 88.3%; Pred. No. 4.4e-31;
Matches 211; Conservative 0; Mismatches 21; Indels 7; Gaps 1;

QY 226 CACGAGAGCGAGCGGTGGAGAGAGGAGGACATGGTCTCCGAGGAGGAGGAGGAG 285
DB 26 CCCGAGAGTACGACGACGACGAGGAGGAGGACATGGTCTCCGAGGAGGAGGAGGAG 85

QY 286 GACGGC-----GACGCCGAGGAGACCCAGGATTCTTGAGACGACGAGGAGGATGAGAT 338
DB 86 GAGGAGGATGGCTGACGCCGAGGAGACCCAGGATTCCGAGGACGAGGAGGAAGATGACAT 145

QY 339 GGAAGAGGAGCAGCATGCTCCGATTATCCGAGGAGGATGAAGACGACGACGACGACG 398
DB 146 GGAAGAGGAGCAGCATGCTCCGATTATCCGAGGAGGATGAAGACGACGACGACGACG 205

QY 399 CAGTTACTGCGGAAAGCAGCTTCAGGAGCCATAGTACCTACAGACGACTCCAGGTA 457
DB 206 TAGTTACTGCGGAAAGCAGCTTCAGGAGCCATAGTACCTACAGACGACTCCAGGTA 264
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 01:09:02 ; Search time 22.9984 Seconds
(without alignments)
5348.092 Million cell updates/sec

Title: US-09-698-295-1
Perfect score: 14971
Sequence: 1 MVSEEEEEGDGAETQDSE.....KLKGFKASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```
Database : Issued_Patents_AA:*
1: /cgn2-6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2-6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2-6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2-6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2-6/ptodata/2/iaa/PTCTUS_COMB.pep:*
6: /cgn2-6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description	
	Score	Match	Length			
1	982.5	6.6	238	4	US-09-257-179-80	Sequence 80, Appl
2	491	3.3	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
3	441.5	2.9	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
4	391	2.6	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
5	386.5	2.6	2136	4	US-09-854-856-14	Sequence 14, Appl
6	380.5	2.5	1999	4	US-09-854-856-16	Sequence 16, Appl
7	380.5	2.5	2108	4	US-09-854-856-30	Sequence 30, Appl
8	380	2.5	3788	4	US-09-914-259-11	Sequence 11, Appl
9	378	2.5	8991	4	US-08-714-741-32	Sequence 32, Appl
10	376	2.5	2076	4	US-09-854-856-46	Sequence 46, Appl
11	374.5	2.5	1971	4	US-09-854-856-32	Sequence 32, Appl
12	374	2.5	2201	4	US-09-854-856-24	Sequence 24, Appl
13	370	2.5	1939	4	US-09-854-856-48	Sequence 48, Appl
14	370	2.5	2048	4	US-09-854-856-62	Sequence 62, Appl
15	369	2.5	2229	4	US-09-854-856-8	Sequence 8, Appl
16	368	2.5	2064	4	US-09-854-856-26	Sequence 26, Appl
17	367	2.5	3969	3	US-08-061-376-5	Sequence 5, Appl
18	364	2.4	1911	4	US-09-854-856-64	Sequence 64, Appl
19	363.5	2.4	2141	4	US-09-854-856-56	Sequence 56, Appl
20	363	2.4	2092	4	US-09-854-856-10	Sequence 10, Appl
21	362.5	2.4	2842	1	US-07-741-940-7	Sequence 7, Appl
22	362.5	2.4	2842	1	US-08-289-548A-7	Sequence 7, Appl
23	362.5	2.4	2842	1	US-08-452-654-7	Sequence 7, Appl
24	362.5	2.4	2842	4	US-08-449-731-7	Sequence 7, Appl
25	362.5	2.4	2843	1	US-08-452-655B-2	Sequence 2, Appl
26	362.5	2.4	2843	1	US-08-452-655B-7	Sequence 7, Appl
27	362.5	2.4	2843	3	US-08-450-592-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

```

US-09-257-179-80
; Sequence 80, Application US/09257179
; Patent No. 6410709
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 29 Human Secreted Proteins
; FILE REFERENCE: P2015p1
; CURRENT APPLICATION NUMBER: US/09/257,179
; CURRENT FILING DATE: 1993-02-25
; EARLIER APPLICATION NUMBER: PCT/YUS98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,073
; EARLIER FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-257-179-80

```

[illegible]

US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 3.3%; Score 491; DB 4; Length 2137;
Best Local Similarity 17.7%; Pred. No. 1.3e-20;
Matches 410; Conservative 357; Mismatches 1006; Indels 546; Gaps 74;

QY 345 ENENEKKIWIYSTKQVLAELIDCLDKDYWEAEL--CKILEEMREIEHRHMDITEDLTNKA 402
DB 12 EXDYKKSLEQQKTRVKIYK----SGKSVKASINIEILLKTMGLFLSKNEIQENVTEKT 67
QY 403 RG-----SNKSFLAAANEIELESIRAKKGDIDNVKSPETEK 439
DB 68 KGHKLKSAKTTALVGGAFTENMLNNHQFAASETPITSEISSNSETVANQSTTIKNS 127
QY 440 DKNETENDSKDAEKNREEDFOSLEKDDKTPDDDPQCKSEEBTEVCDKGNV----S 495
DB 128 QKETVNSTLESNHNSTNKQMSSEVINTDQSE---KAGISQQSSETSNQSELTAYAS 184
QY 496 ANLGDMTNTATSEETSPSGRSPVGCISLTPDSSNNAEKKVASELPQDV-----P 545
DB 185 TDHVESTTNNNTAQDQNKSSNTYKSTQSTQNTSSSEKNISSLNTQSTETKATDSLATS 244
QY 546 EEPNKTCESSNTSATTSTQ-----PNLE-----NSNSELNLSQSBESA 585
DB 245 EARTSTNOISLNTSTSTQSSPTSPANLRFSTFVLNMAAPTSTTTTSSLSNSV 304
QY 586 KAADDPENGERESHTPVSIQEEIVGDFTEKSTGELSESPGAGKASGSTRITILRNPD 645
DB 305 VVKNDNFN----EHNLS-----GSATYDPKGTATLPDA-YSOKGALSINLRL---- 349
QY 646 SKLSQLKSOQVAAAAHEANKLFKEGKEVLVN-----SQGEISRL 685
DB 350 -----DSNRSFRFTGKVLGNRYEGYSPDVTGDDGIGAFSPGLGQI 393
QY 686 STKKEVIMKGNINNYFLQGEKYRYHHQYNTSNFALNKHQHRE-----DH 732
DB 394 GKECAAVGIGLGNAFGF----KLDYHNTSTPKDAKADPRNVGGGAGFACFVSTDR 449
QY 733 D-----KRRHLAHFCLTLPAGE-----FKWNSVHGSKVLITSLTLITQLENNIPSSF 782
DB 450 NGMATTEASAAKLNVOPTDQNSFQDVIDYNGD---TKVMTYIYAGQTFTR----- 497
QY 783 FHPNASHRANWTIKAVOMGSKPREFALALALECAVKPVVMLPIWREFLGHTRLHRMTSI 842
DB 498 -----NLTDWIKN-----SGGTFSLSMSTASTGGAKNLQOVQFGFTEYTESAVAKVRVY 546
QY 843 EREE-KEKVKKKEKQEEETMQOATWXYTPPVKHQV--WKQGEYRVT-----GYG 893
DB 547 DANTGKDIIPPKTITAGEVDATVN-----IDKQLNKLNSGYSYSTDALQNSYS 596
QY 894 GWSWI-----SKTHVRFVFKPLPGNTNVNRYKSLGKTKNNMDENMDSDKRCRSRP 945
DB 597 ETSCTPLKLTNSSQIVYKF-----KDVQGPQISVD-----SQTREVGKTI 638

QY 946 KKIKI-EPDSEKD-----EVKGSDAAKGADQ-----NMEDISKITEKKDQDVKELL 991
DB 639 NPITITTTDNSKQDLVTTTGTGLPSGLSDFDQTTNTITGTPSEVGTTTIT-----VNTTD 691
QY 992 SDSKPKCKEPMEDVDDMKTESHVNCQESSQV-----DVVNVSE 1030
DB 692 ATGNVTSKQFTITITQDTISPVVNTPQSAEVEFTPINPITITATDNGKGVVHTVTGLPQ 751
QY 1031 GFHLRTSVYKKTKSKLDGLLERRIKQFTLEEKORLEKIKLEGKIGKGTSTNSSKNLS 1090
DB 752 GLKFDASTNSIVGTQIGTNTITESTDASGNKTTIKNYE-----VTRNSASDST 803
QY 1091 ESPVITKAKGECQ-----SDSMRQEQSPNANNDOPEDLIQCCOSDSVLRMSDPSHTTN 1145
DB 804 STSIVNSVSTSI NSNSTLSDSVKASQSLSTKSLSL--SASTSNSTSIQASEASTSK 861
QY 1146 KLYPKDRVLDVYSIRSPET-----KCPKQNSIENDIEEKVSDLA---SRQEPKTKSK 1196
DB 862 QL--SESASTSTSDSASEARKSESTSKSTLSLSESTSTSVSDSASVSTSESASTSTSVSG 919
QY 1197 GNDFFIDDSKLASADDIGTLICKNNKPLIQEESDTIVSSKSAHLSSHVPKSTNDRDATPL 1256
DB 920 STSISIDSTSTSTSDSASIKASASTSKLLSESVSTSDSASTSTSVSDSNASTSL 979
QY 1257 SRA-----MDFEGKLGCDSESNSTLENSDTSVSIQDSSEEDMIVQNSESISEQFRREQ 1311
DB 980 SKSTSTSVSDSTSTSDSASTSTSESDSASTS-----LSESTSTSVSDSTSTSTS 1032
QY 1312 DVEVLEPLKCE-----LVSGESTGNCEDRLPVKGTGANGKPKSQOKKLEER 1357
DB 1033 DSASMSASESNKSKSTLSLSESTSTSLSGSTASTSDSASTSTSES--ESDSTSTSLSES 1090
QY 1358 PVNKCSDQIKLNTTDDKNNENRESEKQGT-----STQIN 1395
DB 1091 TSTSLSGSTS--ASTSDSASTSTSESDSTSESTSLSESLSTSVSDSTSASTSEASTSTSE 1149
QY 1396 GKDNPKPIYLLGCELKEISESRVYGVNPEKPVNNINKIIPENDIKSLTVKESAIRPEFNG 1455
DB 1150 SESNASSTSLSGSLSTSI SDSTSTSDSASTSTSES---ESDSTSTSLSEST----- 1199
QY 1456 DVIMEDFNERNSSSTKSHLLSSDAEGNYRDSLETLPSTKESDSTQTTTPSASCPESNSV 1515
DB 1200 STSLSDSTSTSEASTSTSESDSTSE--STSLSESTSTSVSDSTSAST--SDSASTSTSV 1257
QY 1516 NOVEDMEIETSE-----VKVYTS-----SPITSEES-----NLNDFIDENGL 1594
DB 1258 SDSEASTSISELSTSVSDSTSTSDSASTSTSESDSTSESTSLSESLSTSVSDSTSA 1317
QY 1555 PINKNENVN-GESEKRTVITVTMTSTVATESKTIVKEKGDQTVVVSSTENCAKSTVT 1613
DB 1318 STSDSASTSTSESDSASTSLSGSTSTSLSDSTSTSDSASTSTSESDSERASTSLSG 1377
QY 1614 TTTTIVTKLSTPSTGGSVDDIISVKEQSKTVVTVTTVDS---LTTTGGLVTSMVVKYIS 1670
DB 1378 STSTSLSDSTSTSTSDSAS-----TSTSVSDSNASTSLSGSLSTSVSDSTSTS 1426
QY 1671 TRDKVKLMKFRPKKTRSGTALPSYRKFTKTKSI FVLNDDLLKLARKGGIREVPYF 1730
DB 1427 TSDSAS--ASTSESDSERASTSL-----SGSTSTSI-----SDSTSTSDSASTSTSVS 1474
QY 1731 NYNAKPALDIWYPSPRPTFGITWRQTVKSLAGVSLMLRLLWASLRWDDMAAKVPPG 1790
DB 1475 ESN-----TSTSI-----ESLSTSV--- 1491
QY 1791 GGSTRFETSETITTEIIRKRDVGPGYIRPFCYKIRKICPGIVPETPKETTPORKGLR 1850
DB 1492 SDSTSTSDSASTSTSVSDSASTSSSE-----SVSTSDSESTSTSDSAST 1541
QY 1851 SSALRPKRPETPKQTPGVIIETWVAEEELWEIRAFERVEKEKAQAEQAKRLBOQ 1910
DB 1542 STSVSESNSTSLSGST--STSVSDSTSTSTSDSASTSESD-----SDSASTSSSES 1594
QY 1911 KPTVIATSTSTSTSTSTISPAQKVMVAPISGSVTTGTGMVLTATTKVSGPATVTFQKNK 1970

```

Db      1595  VSTSVSDSTASTSESASTSVSDSNSASTSLSESTSTSLSDSTSMET----- 1643
Qy      1971  FPHOTFATWKGQSGNSGVVQQKVLGILIPSTSTGTSOQTFTSFQPTAT-----VTIR 2023
Db      1644  -SDSASTSTESDSDASTSLSDSTSVSESTSTSVSASNSTSTSLSDSRSTSL 1702
Qy      2024  PNTSGS---GGTTNSQVITGPQIRPGMTVIRTPLOQSTLGAIRTPVMVQPGAPQVM 2080
Db      1703  DSTSTSTSEGSTTSE-----SDSDASTSLSEST-----S 1734
Qy      2081  TQIIRGQPVSTAVSAPNTVS---STPGOKSLTSTST- ----NQSSASQPPRPQOQOV 2132
Db      1735  TSIISDSTSTSDSASTSMVSDSNRASTSLSDSTSVSDSTASTSESASTSTRES 1794
Qy      2133  KLTMATQTLQTQGHGGNGGLTVVIGGQGTGQLQILPGQVTVLPQPGQQLMQAAMPNGT 2192
Db      1795  TSASTSLSESTSVSDSTSTSDSASTSTSESDSNSESTSLSESTSVSDST- 1850
Qy      2193  VQRELFPLATTATATSTTTTTVTTTAAGTGEQRQSKLSPQMQRVHQDKTLPPAQSSVGP 2252
Db      1851  -----ASTSASTSTSVSDSNSAST-----SLSGSTSTSVSD 1883
Qy      2253  AKAPQTAQPSARQPQTQPSRPAQVEVQTQTVTSSHYVPSEAOPTHAQSKKQV 2312
Db      1884  STSTSTASASTSTSESDSDASTSLSGSTSTSDSTSTSDSASTSTSESASTSTSV 1943
Qy      2313  AASQPSQNVGQSGPVVRQSPSQT----RIRPSTPSQLSPGQSQVQVTTTQPIPIQPH 2368
Db      1944  SEDSESTSVSESTSVSDSTSTSTSESASTSTSESESTSESTSVSESTSTSDSSST 2003
Qy      2369  SLQIPSGQPSQPSQVQVQSTQTLSSGQTLNQGVSPPSR 2407
Db      2004  STSMSTSETFTSQSPINSESQFI--GDSLSEDTIVTQSK 2040

RESULT 3
US-09-134-001C-5080
; Sequence 5080, Application US/091344001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match      2.98; Score 441.5; DB 4; Length 3696;
Best Local Similarity 17.1%; Pred. No. 2.7e-17;
Matches 471; Conservative 470; Mismatches 1066; Indels 749; Gaps 106;

Qy      311  QKNKPYIRHPGIGVD-RSRKKYWFNLNR--LLIIEEDTENENEKKIWIYYSKVLAEIDC 367
Db      569  QVAKDII---PSNTLYASYNKYKLEKRAQTVLDEETNTP----- 606
Qy      368  LDKDYWEAEALCKLLEEMREETHRMDITEDLTLNKARGSNKSFLLAAANEELIESIRAKGD 427
Db      607  FNQRYSTQIDDLHLHELQTLINRVSAAREINDKA-----QEMTDAV----- 648
Qy      428  IDNVKSPETEKKNETENDSKAEKNREEFDGSLKSDDKTDPDDPEOGK----- 480
Db      649  -----YDSTELTTEKQFLVQIENHKNKWEISNNIDDELTDGQVERVKAGLHTL 697

```

Qy	481	-SEEPTEV-----GDGRGNSYSANLGDNTTNTATSE-----ETSPSEGRSPVGC	522
Db	698	ESDTPHVPKPNARQVVNNRADQOKTLIRNHEATTEQNEAIRQVEAHSSDAIAKIG-E	756
Qy	523	SETPDSSNMAEKKVASLPDVPPEPKNKTCESSNTSATTTTSIQPNLENSNSSSE-----	576
Db	757	AETDTTVNEARDNGTKLIATDV---PNPTKKAARAAVNTSANSKIKDINNNTQATLDER	813
Qy	577	-----LNSOSE-----SAKAADD---PENG-----ERESHPTPSIOEEIV-----	609
Db	814	NDALALVNRSKDEAIQININTAQNDDVTEAQNNGTNTIQOVLTPVKRQNAIATINAKAD	873
Qy	610	-----GDFTEKSTGSELSESPGAGKSGSGSTRITIRLRNPDSKLSQLKSQOQVAAA	659
Db	874	EQRLIQANNATTEEKADAERKVNNAV---ITANQNITNATTNRVDQAOTGTSGGIISA	930
Qy	660	AHANKLFKEKG-----EVLVYNSQGEISRLSTKKEVIMKGNNTNNTFKLQGEKRYVYHN	714
Db	931	ISPATKTEKEDARAAVEAKAIAQOQINSNNMATTEE---KEDALNQVEAHKQAAIATINQ	987
Qy	715	QYSTNSFALNK-----HOHRBDHKRRHLAHKFCCLTPAGEFKWNGSVHGSKVLJTISTL	767
Db	988	AQSTQOVSEAKNGINTINDQPNVAKKNT---KTILEQKNGEK-KSAIAQTDPDATTEEK	1044
Qy	768	RLTITOLENNIPSPFFHPNNAWASHRANIKAVQCMCKPREPALALALECAVKPVMLPIW	827
Db	1045	QEAVSAVSQAVTNGIITHINQANSND---VDOELSNAEQIITHTNVNVQKPKQA	1095
Qy	828	REFIGHTRLHRMTSIEREEKEVKKBEK-----KQEEETMQOATNVKYTFPVKHQV	879
Db	1096	RQALIAKTNEKQAINSDNEGTTTEEKQAQISLNDAKNLADEQITQAA-----	1143
Qy	880	WKQGEERYVYGGGWSWISKTHYREVPLPGNTNVNRYKSLCTKNNNDENMDESDDR	939
Db	1144	---SNQVNDNALIGINSIKIOT-NFTKKQARDQVNOQFQBEAKELNSTPHATODEKQ	1199
Qy	940	KC---SRSPKKIIEPDSEKDEYKSGDAAGAQDNEMDISKITEKKDQDVKELLDSDSK	996
Db	1200	DALTRLTOAKETALNDINOATNQNVDTALTSGIQNIQNTQVNVRRKQEAKTII-NDIVQ	1258
Qy	997	PKCEEPMEVDDDMKTESHV--NQESSQVDVN-----VSEG-----FHL	1034
Db	1259	QHKQSIQNDDDATTEKEVANLVNASQQNVVISKIDNATTNNQIDGIVSDGRQSIINAITP	1318
Qy	1035	RTSYK-----KTKSKSLDGLLE-----RRIKOFTLEEKORLEKIKLEGG	1074
Db	1319	DTSIKRNAKNDIDIKAADKKIKITORINDAATDEFIQEBANRKIEAKTEAKTIONSTROQ	1378
Qy	1075	IKGIGTSTSNKSLNESPVI-TPAKEGQCSDMRO---EQSPNANNODPEDLIOGCSQ	1129
Db	1379	VNEAKTNGINKINIPATTVKSEARQAVQNKANEQINHINTQNPDTATNEKQEAIRVSA	1438
Qy	1130	SDSSVLRMSDPSSHNTKLYPKDRVLDD--VSIRSPETKCPQNSIENDEEKVS-----	1181
Db	1439	ELARVOAQINAEHTTQGV---KTIKDDAITSLSRINAAVVEKESARNAIEQKATQOTOFI	1495
Qy	1182	-----DLASRQOETPKSK-----TKGDNDFDIDSKLASADDIGTLI-----	1217
Db	1496	NNNDNATDEBEYKANLVIAITAKSLDNLINSLSNND--VENAKVAGINEANVLPATAV	1553
Qy	1218	-CKNKKPLIQEESDVTVSSSKSALHSVPKSTNDROT-PLSRAMDFEGKIGCDSE--SN	1273
Db	1554	KSAKKDIDOKLAQI-----NOIQTHQATTTTEKEAAIQLANQKNEARTATQNEHSNN	1608
Qy	1274	STLENSSDTVS-----IQDSSEEDIMVQNSNEISEQFR-----TREQDYEVLEPLK	1320
Db	1609	GVAQAKSNGHIEIYVMPDAHKKSDAKQSIDKNYNEQSNITNTTPDATDEEKQKALDKLK	1668
Qy	1321	CELVSG-----ESTGNCERDLRPLVKGTGAN-GKKPQOQKLEER-----	1357
Db	1669	IARDAQYKQVDAQTNOQVSDAKTEAIDTTTNTQANVAKPKPSARVELDKSFKEDLKRQINA	1728

1358 QY -----PVNKCSDQIKLWNTDKNNENRESEKK-QRTSTFOI----- 1394
1729 Db TPNAETEERKQDAIQLRNGKRDVKNLQNRDRDNEVEHQHKNIGLOELETTHANPTRKSDA 1788
1395 QY -----NCKDNKPKIYLGEGCLKEISERVSNGVNEPKVNN----- 1429
1789 Db LOELQTKFISQTELINNNKDATNEEKDEAKRLLEISKNNKTITNINOAGTNNOVDNAKDMG 1848
1430 QY -----INKIIPENDIK-----SLTVKESAIRPFING--DVIMED-----FNERNSESTKSH 1473
1849 Db MNEIATIPATIKTDAITDKKAEQOQVTIINGNNDATDEKAEARKLVEKAKIEAKSN 1908
1474 QY LLSSS---DAENGYRDSLETLPSTKESDSTQTPSPASCPSNSVNOVEDMEIETSEVKK 1530
1909 Db ITNSDTEREVNGAKTNGLEKINNIQIPSTQTKT-----NAKQELNDRKAEQ---QLIQ 1956
1531 QY VTSSPITSSEENSLNDFIDENGLPINKNNVNGESKRKTIVITEVTMTSTVATESKTVI 1590
1957 Db INNTPDATEEKQEAATRVNA-GL-AQAIONINNAHSTQEVNESKTSIATIKSVOPNVI 2014
1591 QY K-----VEKGDKQTVVSTEN-----CAKSTVT-TTITVTKLSTPSTGSGVDI 1633
2015 Db KKPTAINSLTOEANNQKTLIGNDNATDEKAEAKOLVTQKLINEOIQIHSTQDNQVD- 2073
1634 QY ISVKEOSKVTWTTVTDS-----LTTTGGTILVTSMVTSKEYSTRDKVKLMKFSRPK 1684
2074 Db NVKAGIITAILINANAHRQDAINILTNLAESKSDIRANODATTEK----- 2122
1685 QY KTRSGTALPSYRKFTVKSTKSFIVLPDDDLKKLARKGIREVPYFNNAKPAIDWYP 1744
2123 Db NTAIQSI-----DNTLAQAR-----NNINGANTNA----- 2147
1745 QY SPRPTGITHRYLQTVKSLAGVSLMLRLMLASLRWDDMAAKVPPGGSGSTRTSETET 1804
2148 Db -----LVDENLEDGKQLQRIVLSTQTKQAKADIAQAIGQORSTIQNQNA 2194
1805 QY TT-----EIKR-----RDVPGYGI---RFEYCI 1825
2195 Db TTEKQEAERLQNETNGVNDRIOAALANQNVTDENKNILETIRNVEPIVIVKPRANEII 2254
1926 QY RKIICPGVETPKETPTQPKGLRSALRPRPEPKQTPGVPIETWVAEELEWEIR 1885
2255 Db RKAAEQTTLINQNDATLEEKQIALGLKEEVNEALNQ-----VSAHNNNDVKIAENN 2309
1886 QY AFARVEKKAQAQVEOAKKRLQKQKPTIATSTTSPTSTSTISPAQKVMVAPIS--G 1943
2310 Db GIAKISEVHPETIKRNAKOEIEQDAQSQDITNANNKSINEEKSAAIDRVNVAKIDAIN 2369
1944 QY SVTTGTKMWLTGVGSPATVTFQONKNPHQTFATWVKQGSNGVGVQVQKVLGIIPSS- 2002
2370 Db NITNAT-----TTQLVNDA-----KNSGNTSISQ-----ILPSTA 2399
2003 QY --TGTSQOTFTSFQPRATVITIRPNTSGSGTTSNSOVITGPQIRPGMTVIRPLOQ--- 2057
2400 Db VKTNALAALASEAKNNWAIIDQTPNATAEKEEANKV-----DRLQEAD 2445
2058 QY STLKGAIRTPV-MVOPGAPQQ---VMTQIIRQGPVSTAVSAPNTVSSTPGQKSLSAT 2113
2446 Db ANILKAHTDDEVNNIKQAQVQINAVQVEVIKKQNVKNQLN-----QFIDNOKKIIENTP 2500
2114 QY TSNTOSSASQPPRQOQGVKLTMAQLTQLTQGGHNGOGLVIVTQGGQTTGQLQLPOGV 2173
2501 Db DATEEKA-EANRLQNVLTSTDEIANVDHNNVEVDQALD-----KARPKIE 2546
2174 QY TVLPCGQOLMQAAMPNGTVQREFLTPPLATTATTSTTTVTSTTAAAGTGEQRQSKLSPQ 2233
2547 Db EIVP---QVSKRRDVLNAIQEAF-----NSQTOEIENQENATNEEKEALNKI 2591
2234 QY MQ-VHQDK-TLPPAQSS--SVGPAKAQPQTAQPSARPQPTQPOSPAPQEVQTOPEVQTOT 2290
2592 Db NOLLNQAKVNDIQAGSNKDVDSAKTRTSIQDIEQIOPHPQTKATGRHRLNEKAN---QQQS 2648
2291 QY TVSSHVPSEAOPTHAQSKKPQ-----VAAQSQPSN-----VQGSPPRVQSPSQTR 2337

2649 Db TIATHPNSTTEERQEAQAKLQAEVLKKAIAKIDRGOTNDDVEKTVVNGIAEINILPA-TT 2707
2338 QY IRSTPQSLSFGQSQVOTTTSPQIPQIOPHTSILQPSQSQPOSOQVQOSTQTLSSGQTL 2397
2708 Db VKDKAKADVNAEKEQ-----NQIINSNDTEATTEELKVLASDNLNHVVET 2752
2398 QY NOVSVSPSRPQLQIQP-----QPOVIAVFPQLQOQVQVLSQISQOVVAAQIAOQS 2448
2753 Db NQAIEDAPDNTNOVNVEKNKGIGITRDIQLPLVKKPTAK-----SKIESAV----- 2797
2449 QY GVPQIKLQPLQIQSSAVQTHQIQNVVTVVAAASVOEQQLQVQOOL-----RQOQKKKQ 2503
2798 Db -----EKKKTEINQTONATHDE---VREGLNQLNQTHEKAKNDVNOQSTN 2839
2504 QY QQIEINVTPSKLLIKVEIILQKVVMKHNAV-----IEHLKQKKSMTFAERE--- 2550
2840 Db QQVE---NAEONSUDQINFRPDPFSKARNAVAEIVKAAQNKIDIEIQEFSATQEKDNAL 2896
2551 QY -----ENQRMIVCNQVMKY-----ILDKIDKEKQAA 2577
2897 Db QHLDEQVKEIINSINQANTDNEVDNKTSGLNITTEYRPEYNNKKNAILKLYDVSDTQEA 2956
2578 QY KKRREESVEOKRSQONATKLSALLFKHKEQLRAEILKRALLDKDLQIEVQBELK 2633
2957 Db IINGYPDATEDELOQAN-SKLNKILLDAKQOIGLAHTNNEV---DDIYNEVSQMKM 3008

RESULT 4
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3159

Query Match 2.6%; Score 391; DB 4; Length 10182;
Best Local Similarity 16.9%; Pred. No. 1,3e-13;
Matches 500; Conservative 450; Mismatches 1114; Indels 894; Gaps 116;

QY 287 EDEWQCEVCAHKVPGVTDCAVEIQKNKPYIRHEPIGYDRS-----RRKYWFLNRRLLI 340
Db KQQLQSIQPAASTDGTODSTRNKKNRQAAEQAIQANSVINNGDATSQINDAKNTV 2725
QY 341 EEDTENENKKTWYTSKVLAEIDCLDKDYW-----EALCKILLEEMREEIHRHMDITE 396
Db EQAQRDYVEAKSNLRADKSQLQSDYDNLNRDLTNDKPPASVRRYNEAISNTRKELDTAK 2785
QY 397 -DLTNKARGSNKSF-----LAAANEELIESIRAKKGDIDNVKS 433
Db ADASSTLRNTNPSVEQVRDALNKINTQPKVNOQATLQPKNNSELVQAKRLQDAVND 2845
QY 434 PETEKDKNETENDSKDAEKNEE---FEDQLEKSDDKTPDDDPQKSEPTVEV--- 487
Db IPTQGTQGTQOTINNYNDKQREARALTSQARVDINGDATTOEITSEKSKVEQAMQALTNA 2905
QY 488 -----GDKG-----NSVSANLGDNTTATSEETSPSEGRSPVGCSETPSSNMA 532
Db KSNLRADKNELOFATYAKNLIENVSTN-GKKPASIRQVETAKARIONIN-----DAKNEA 2958

Db 991 PQQIGIPT-----SSLTQV-----VHSAGRRFIVSPVPESKRE 1024
QY 1421 GNVEPKVNNINKIIPENDIKSLVAKESAIRPFINGDVMEDFNERNSSETKSHLLSSDA 1480
Db 1025 SKVFP-----SEITDTVAASTAQSPGN---LSHSASSLSLQQAFASELFRAQMT 1070
QY 1481 EG-----NYRDSLETLPTK---ESDSTOTTPPSACPSNSVNOVEDMEIETSEVKV 1531
Db 1071 EGPNTAPPNFSHTGPTFPVPPFLSLSTAGVPTTAATAPVATSPSPND--ISTS-----V 1124
QY 1532 TSSPITSEESNL-----SNDFIDENGLPKNNENNGESKRTVITEVTMTSTVATES 1586
Db 1125 IQSEVTPTEGATGATSTGVTSGLPI---PPVSESPVLSVSSVSSITIPAVVSIIST 1181
QY 1587 KTVIKVKGDKQKQVWSSTENCAKSTVTTTTTKLSTPSTGGSDVIISVKEQSKTVVTT 1646
Db 1182 SPSLOVPTSTSEIWSSTALVPSVTSATSA-----SAGGSTATPGPKPPA--VVSQ 1231
QY 1647 TVTDSLTTTGTGLVTSMTVSKEYSTRDKVKLMKFRPKKTRSGTALPSYKRFYTKSKKS 1706
Db 1232 QAAGS--TTVGATLTSVSTTTSFSTASQLSI-----QLSSSTSTPTLAETVVSAAH-- 1281
QY 1707 IFVLPNDLKLARKGIREVPYFNYNNAKPAIDWPVPSRPRTFGIWRVRLQTVKSL-- 1764
Db 1282 -----SLDKTSHS---TTGLAFSLSAFSSSSSPG 1308
QY 1765 AGVSLMLRLWASLRWMDMAAKVPPGGGSTRTETSETETTTTETIIRRDVGPYIRFEYC 1824
Db 1309 AGVSSVIS-----QPGG----- 1320
QY 1825 IRKLIICPIGYPETPKETP--TPQRKGLBSALRKRKRPETPKQTGPV-----IETWVAEE- 1877
Db 1321 ----LHPLVIPSVIATPILPQAAGPTSTPLLPQVPSIPPLVQVPAVPAVQOVLTHSOP 1376
QY 1878 -----ELE-----LWEIRAFERVE---KEKAQAEQOAKRLEEQ- 1910
Db 1377 OPALLPNQPHTHCPEVDSDTOPKAPGIDDIKTLEELKRLSLFSEHSSGAQHASVLSLTS 1436
QY 1911 --KPTV---IATSTTSP-----SSTTSTISPAKV---MVAPISGSVTTGKMLVTKVG 1958
Db 1437 VIESTVPGIPTTAVAPSCLLTSTTCLPPTPLPLGTVALPVPVTPG-----QVS 1489
QY 1959 SPATVFOQKNHFQFATVWKQGSNGVGVVQVQKVLGIIPSTGTSQOTFTSFOPRTA 2018
Db 1490 TPVSTT-----TSQVKGPGTAPSKPLTK 1512
QY 2019 TVTIRPNTSGGTTNSOVIPTGPIRPGMTVIRTPLOQ-----STLG 2061
Db 1513 APVLPVGTETLPAGTLPEQLPFP--GPSLTQSQOQPLEDLDAQLRRLTSPXETVTVTSVG 1570
QY 2062 KAIIRTPVMVQGA--PQOVMTOIRGQPVSTAVSA-----PNTVSTPGQKS--- 2107
Db 1571 PVSMAAPTATTEAGTQPKGVSOVKEGVLATSSGAGVFKMGFRQVSVAAADGAQKEGKN 1630
QY 2108 -----LTSATSTNIQSSASOPRPOGOVKLTHMAQLTQLTOGHGNOGLTVVIQGO 2159
Db 1631 SEDAKSVHFESSSTSSVLSSSS-----PESTLVKPE----- 1662
QY 2160 GOTTGQLIIPQGVTVLPFGGOQLMOAAMPNGTVQRFLETPPLATTATTASTTT----- 2212
Db 1663 -----PNCITI--PGISSDVPESAH-----KTTASEAKSDGQPTKVG 1698
QY 2213 -TTVSTTAAGTGBQROSK-----LSPQMQVHQDKTLPPAQSSSYGPAKAQP 2257
Db 1699 RFQVTTTANKVGRFSVSKTEDKITDKKEGPVASPPFMDLEQAVLP-----AVIPKKEP 1753
QY 2258 QTAQPSARPOPOQPO-----SPAOP----- 2278
Db 1754 ELSEPSHLNGPSDPAEFLSRDVGSGSPHQLSSKSLPSQNLSSLSNFSNSYM 1813
QY 2279 -----EVO-----TOPEVOTQTVSSHVP----- 2297
Db 1814 SSDNESDIEDDLKLELRRLRDLKHLKEIQDLQSRQKHEIESLTKLGKVPVPAVLIIPPAAP 1873

QY 2298 ---SEAQTHAQSSKQVAA---OSQPQ---SNVQSQSPVRVQSPQSTRIRFPSTPSQLSPG 2349
Db 1874 LSGRRRPTKSKGSKSSSSGLNKNKSPQLSGNLSQSAASVHLHPQQT-LHP--PGNIPES 1930
QY 2350 QQSQVQTTTQOPPIQPHT-----SLQTPSQGQSQSQSQSQSQSQSQSQSQSQSQSQ 2398
Db 1931 GONQL-----LQPKPSPSSONLYSAFTSDGAISVPSLSAP---CQGTSSNTV--GATVN 1981
QY 2399 QVSVSSPSRPQLQIQOQPOP 2417
Db 1982 S-----QAAQAP 1989

RESULT 6
US-09-854-856-16
; Sequence 16, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Waite, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; PRIOR FILING DATE: 2001-05-14
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1999
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1999)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-16

Query Match 2.5%; Score 380.5; DB 4; Length 1999;
Best Local Similarity 18.0%; Pred. No. 5.2e-14;
Matches 434; Conservative 297; Mismatches 757; Indels 921; Gaps 104;

QY 472 PDDDEQKSEETEVGDKNGSVSANLGDNATNATSE-----ETSPSEGRSPVCLSETP 526
Db 21 PAPAPKNGSSD--SSVGEK---LGAADAADVTGTEEYRRRRHTMDKDSGAAATTTTTE 76
QY 527 -----DSS-----NMAEKKVASLPQDPPEPNKTCESNTSATTTSI- 564
Db 77 HRFRRSVICDSNATALPGLPLSLPQSPIPAAPVQSAPEPHR--EETVTATATSOVA 134
QY 565 -QPNLENSNSSELNSQSESAKAADPEN--GERESHTPVSTQEEIVGDTSEKSTGEL 621
Db 135 QQP-----PAAAPGEQAVAGPAPSTVPSTSKDRPVS-QPSLV-----GSK 175
QY 622 SESPGAGKAGSGSTRIITRLRNPDLSQLSKSQVAAAHEANKLFREGKVLVNSQGE 681
Db 176 EEPPEARSGGGG-----AKEQEERSQ-----QDD 203
QY 682 ISRLSTKKEVIMKGNINNYFKLQGE---GKRYVHNOYSTNSFA-----LN 724
Db 204 IELETK--AVGNSNDGRFLKFDIEIGRGSFKTVYKGLDTETTVAVAWCELQDRKLTKE 261
QY 725 KQHREDHDKRRHLAH-----KFCLETPAGEFKWNGS-----VHGSKVLTI 764
Db 262 RQRKEAEAEMLKQLQHPNIVRFYDWSWESTVKGKCKIVLTELMTSGTLTKYLRKVKMKI 321
QY 765 STLRLTITOLENIPSPFFHPNW-----ASHRANW 794
Db 322 KVLRSWCRQKLKGL--QFLHTRTPPIIHRDLKCDNIFITGTGTVSKVIGDLGLATLKRAFS 379

APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
FILE OF INVENTION: Encoding the Same
FILE REFERENCE: LEX-0178-USA
CURRENT APPLICATION NUMBER: US/09/854,856
CURRENT FILING DATE: 2001-05-14
PRIORITY APPLICATION NUMBER: US 60/206,015
PRIORITY FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 64
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 2108
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(2108)
OTHER INFORMATION: Xaa - Any Amino Acid
US-09-854-856-30

Query Match 2.5%; Score 380.5; DB 4; Length 2108;
Best Local Similarity 17.9%; Pred. No. 5.6e-14;
Matches 430; Conservative 300; Mismatches 751; Indels 925; Gaps 105;

QY 472 PDDDPQGGSEETEVGDGKNSVANSALGDNNTNATSE-----ETSPSEGRSPVGCLESTP 526
D 1 PAPAPKNGSSSD--SSVGEK---LGAADAAVATGRTTEYYRRRRHTMDKSRGAAATTTTTE 76
QY 527 -----DSS-----NMAEKVASELPQDYPEEPNKTCESSNTSATTSI- 564
D 77 HRFRRSVICDSNATLEPLGLPLSLPQSPAPVQSPAPPEPHR--EETVTATATASQA 134
QY 565 -QPNLENSSSSLSNQSSESAAKADDPEN--GERESHTPVSTQEEIVGDFTEKSTGEL 621
D 135 QQP-----PAAAPGEQAVAGPAPSTVPSTSKDRPV-SPSLV-----GSK 175
QY 622 SESFGAGKASGSTRITRLRPNDPSKLSQKSOQVAAAHAANKLPKEGKVLVNSQGE 681
D 176 EEPFPARSGGGS-----AKEPQERSQ-----QQDD 203
QY 682 ISRLSTKKEVIMKGNINNYFKLQGE--GKRYVHYNOYSTNSFA-----LN 724
D 204 IELETK--AVGMSNDGRFLKFDIEIGRSGFKVYVGLDFTTVEVAVCELQDRKLTKE 261
QY 725 KHQHRDHDKRRHLAH-----KFCLTTPAGEFKWGS-----VHGSKVLTI 764
D 262 RQRFKEEAEMKGLQHPNIVRFYDSWESTVKGKCIVLVTELMTSGTLKYLKRFKVMKI 321
QY 765 STLRLITQLENNIPSPFHPNW-----ASHRANW 794
D 322 KVLRSWCRLKGL--QFLHTRTPPIIHRDLKCDNIFITGPTGSGVIGDLGLATLKRASF 379
QY 795 IKAV-----QMSKPRE-----FALALATLECA-----V 818
D 380 AKSVIGTPEMAPEMEYKDESVDYAFGCMLEMATSEYPYSECONAAQIYRVTSV 439
QY 819 KPV-----VMLPIWRE-----FLGHTRLHRMTSI-----EREKEK-- 849
D 440 KPASFQKVAIPEVKEIIEGCIQKNDERYSIKDLNHPFQEEYGVYVELAEDDEGKTA 499
QY 850 -----VKKEKKOEETEMQQAATWVYTPPVKHQVWKKQGEYRVYTG---GGWSW 897
D 500 IKLWLRIEDIKLKGYKQNEA-----IEFSPDLERDVPEDVAQEMVESGYVCEGDHKT 553
QY 898 ISKTHYVRFVPLPGNTNRYKSLGKTKNNMDENDESCKRCSRPKKIKLEPSEKD 957
D 554 MAK-----IKDRVSLIKRKEQQLVREOEKK-----KQEESSLKQ 591
QY 958 EVKGSAAKAGADQENMDISKITEKQDQVKLELDSDSKPCKEPMEVDDMDKTESHVNC 1017
D 592 QVEQSSASQ-----TGKQLPSASTGIP-----TASTTSA 621
QY 1018 QESSQVDVNVNSEGFHLRTSYKKKTKSKLDGLLERRIKOFTLEEKORLEKIKLEGGIKG 1077

Db 622 SVSTQVEPEPEADQHQLOYOQPSISVLSDGTVD----- 656
QY 1078 IGTSTNSSKNLSESPVITKAKGECQSDSRQEQSPNANDQPEDLIQGCQSQSDSYLRM 1137
Db 657 ----SGQSSVFTESRV-----SSQQVSYSGSQHQAHSHTVPGHPTSTVQAQ 701
QY 1138 SDPSHTNKLKYPKDRVLDDVYSIRSPETKCPQKNSIENDI-----EEKYSDLASRQEP 1190
D 702 SQP-----HGVYPPSSVQOGI-----QQTAPQOQTVQVSLQSTSTSSSEATTAQPVSPQAP 752
QY 1191 T-----KSKTKGNDFFIDDKSLASADDIGTLCKNKKPLQEEESDITVSSSKSALHS 1242
D 753 QVLPQVSAGKQSTQG-----VSQVAPAEPAV-----AQP--QATQPTTLASSVDSAHS 799
QY 1243 SVPKSTNDRDATPLSRAMDFEGK-----LGCDSNSSTL 1276
D 800 DVASGMSDGNENVPSSSGRHEGRTTKRHYKRSVRSRHEKTSRPKRLILNVSNKGRVV 859
QY 1277 ENSSDT-----VSIQ-----DSSEE-----DMIVONSNEISISEQFRTREQDVEYLEP 1318
D 860 ECQLETHNRKMWTFKFDLDGDNPEEIATIMVNDFFILAIERESFVDQVR---EIEKADE 916
QY 1319 LKCELYSGESTGNCEDRLPVKGTANGKPKSQOKKLEER---PVNKCSDQIKLKNTTDDK 1375
D 917 MLESDVSEPEGD--QGLESLQGDYDYGSGOKLEGEFKQIPASSMPOQIGIPT----- 970
QY 1376 NENRESEKKQRTSTFQINGKDNKPKIYLKGE--CLKEISESRVSGNVEPKVNNINKI 1433
D 971 -----SSLTQV-----VHSAGRFFVSPVPSKRLRESKVPF----- 1001
QY 1434 IPENDIKSLTVKESAIRPFINGDIVMEDFNERNSSSETKSHLLSSDAEG-----NYRDS 1487
D 1002 ---SEITDVTAASTAQSPGN---LHSASSLSLQQAFLSELRAQMTGEPNTAPPNESH 1055
QY 1488 LETLPSTK---ESDSTQTTTPSACPSNSVNOVEDMEITSEVKKVTSSPITSEESNL 1544
D 1056 GPTFPVVPPLSLSIAGVPTTAAATAPVPATSSPND--ISTS---VIOSEVTVPTTEGI 1109
QY 1545 ---SNDFLIDENGLPINKENNVNGESKRKTVTTEVTMTSTVATESKTVKVEKGDKOT 1599
D 1110 AGVATSGVTVSGGLPI---PPVSESPVLSVVSSVITIPAVSISTSPSLQVPTSEI 1166
QY 1600 VVSTENCAKSTVTTTTVTKLSTPSTGGSDIISVKEQSKTVVTVTTVDSLTTTGGTL 1659
D 1167 VVSSTALYPSVTVSATSA-----SAGGSTATPGPKPPA--VVSQAAAGS--TTVCATL 1215
QY 1660 VTSMTVSKYSTDKVKLMKFSRPKTRSGTALPSYKFKVTSTKKSIFVLPNDLKLKLA 1719
D 1216 TSVSTTTSFPSTASQLSI-----QLSSSTSTPTLAETVVSAAH----- 1253
QY 1720 RKGIREVPYFNNAKPAIDWIPYSPRPTFGITWRYLQTVKSL--AGVSLMLRLLAS 1777
D 1254 -----SLDKTSHSS---TTGLAFSLASPSSSSPGAGVSSYIS----- 1288
QY 1778 LRWDDMAAKVPPGGGSTRTETSETIETTBIIRRDVGPYGIREFEYCIKRIICPIGVET 1837
D 1289 -----QPGG-----LHPLVIPS 1301
QY 1838 PKETP--TPQKRGRLRSALRPKRPETPKQTPV-----IETWVAEE----- 1877
D 1302 IASTPILPQAAGTSTPPLPQVPSIPPLVQVAVNPAVQOQTLIHSOPQAPALLNPQPHTC 1361
QY 1878 -ELE-----LWEIRAFARVE---KEKAQAVEQOAKRLEQ---KPTV---IATS 1918
D 1362 PEVDSQTQPKAPGIDDIKLTLEKLSLSESHSSGAHASVLSLTVIESTVTPGIPPT 1421
QY 1919 TTSPT---SSTTSTISPAQKV---MVAPISGVTTCMKVLTTKVGSPTATVTFQOKNF 1971
D 1422 AVAPSKLLTSTTCLPPTNLPLGTVALPVPVTPG-----QVSTPVTST----- 1467
QY 1972 HQTATVWKGQSGNSGVVQVKVLGIIPSTSTQSTFTSFQPRATVATVIRPNWTSGGG 2031

Db	1468	-----TSGVKPGTAPSKPPLTKAPVLPVGTLPAG	1497
Qy	2032	TTNSOVITGQIRPGMTVIRTPLOQ-----STLGKAIIRFPVMVOPG	2074
Db	1498	TLPSPQLPFP--GPSLTQSOQPLEDLAQLRRLTSPXITVTSVAVGVSMAAPTATEA	1555
Qy	2075	A--POOVMTQIRGPVSTAVSA-----PNTVSTPQGS-----LPSAT	2112
Db	1556	GTQPGKGVSVQKEGVPVLTATSGAGVFKMGFRQVVAADGAGQKEGKNKSEDAKS	1615
Qy	2113	STNSIQSSASQPPRPOQGOVQLTMAOLTQLTQGHGNOGLTVTIQOGQITGQLQLIPQ	2172
Db	1616	SESSLSSSS-----PESTLVPE-----PNG	1637
Qy	2173	VTVLPQGOQLMQAAMPNGTVQRFLETPLATATTAFTTT-----TTVSTTAAGTGE	2224
Db	1638	ITFI-PGSSDVESA-----KTTASEAKSDTQPTKVGRFQVTTTANKVGR	1683
Qy	2225	QRQSK-----LSPOMQVHQDKTLPAPQSSSVGPAKAPQTAOPSARPOT	2270
Db	1684	FVSKTEKIDTKKEGVASPPFMDLEQAVLP-----AVIPKKEKPELSEPSHLNGPSS	1738
Qy	2271	QPO-----SPAQP-----	2278
Db	1739	DEAAFLSRVDGSGSPHSQLSKSLPQNLSQLSLSNSFSNSYSMSDDNESDIEDL	1798
Qy	2279	-----EVO-----TQPEVQQTQTVSSHVP-----SEAQPTHAQS	2307
Db	1799	KLELRRLDKHKEIQDLSROKHEIESLYTKLGVPPAVIIPPAPLSGRRRRPTKSG	1858
Qy	2308	SKPQVAAA--QSQPO--SNVOGQSPVRVQSPSOTRIRPSTPSQLSPGQSOVQTTTSQPI	2362
Db	1859	SKSRSSSLGNKSPQLSGNLSCQSAASVLPQOT-LHP--PGNIPESQNL-----LQPL	1911
Qy	2363	PIQPH-----SLQIPSGQPOQPOVQSGSTOTLSGQTLNQVSSVSPRPOLQ	2411
Db	1912	KPSPSDNLYSAFTSDGALSVPLSAP-----GQGTSTNTV--GATVNS-----Q	1955
Qy	2412	IQOPOP	2417
Db	1956	AAQAQP	1961
RESULT 8			
US-09-914-259-11			
; Sequence 11, Application US/09914259			
; Patent No. 6495336			
; GENERAL INFORMATION:			
; APPLICANT: Makowski, Lee			
; APPLICANT: Hyman, Paul			
; APPLICANT: Williams, Mark			
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES			
; FILE REFERENCE: 8471-010-999			
; CURRENT APPLICATION NUMBER: US/09/914,259			
; CURRENT FILING DATE: 2000-11-21			
; NUMBER OF SEQ ID NOS: 1800			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 11			
; LENGTH: 3878			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-914-259-11			
Query Match 2.5%; Score 380; DB 4; Length 3878;			
Best Local Similarity 16.8%; Pred. No. 1.4e-13;			
Matches 488; Conservative 476; Mismatches 1025; Indels 922; Gaps 115;			
Qy	305	DCVAEI-----OKNKP--YIRHEPIGYDRSRKWK-----FLNRRLLIE	341
Db	125	DCSEVNGCSFVMRTGKPTNLLRREEFVDDSYSEQAQSPTHLEMMSELAGKQHEIE	184
Qy	342	EDTENEKKIWIYSTKVQLAELIDCL-----DKDYWEAELCKIL	381

Db	185	ELNRELEMRVYTGTEGLOQLOFEAAIKORGIITQLTANLQQAARREKDETMREFLELT	244
Qy	382	EMRE-EIH-RHMDITEDLTNKARGSNKSFLLAANEIEILSIRAKKGGIDNNKSPETEK	439
Db	245	EQSOKLQIQOQLOASTLNRNSTHSTAAADLQAKQOILTH-----QQ	287
Qy	440	DKNETENDSKAENREFEFQSLKSDDKTDDPDEQKSEEPTEVGDGKNSVANLG	499
Db	288	QLEQDHLLEDYQKKKEDFTMQI--SFLQEKIKVYEMEQDKVE-----329	
Qy	500	DNTNATSETSPSEGRSPVCLSETPDDSNMAEKVASELPDVPPEEPNKTCESNTSA	559
Db	330	---NSNKEIEQKE-----TIEELNTKIIIEEKKTLELKDCLT	365
Qy	560	TTTS-----IQPNLENSSSSELNSQSSAKAADP-----PENGERSHTPSV	603
Db	366	TAKLLGELOEQIVQKQOEIKNNKLELTNSKQKERSQSEIKQLMGTVELOKRNHKDSQ	425
Qy	604	IQEIVGDTSEKSTGSELSPGAGSGSTRITITRLRNPDSKLSQKQVAAAHEA	663
Db	426	FETDIVQMEQE-----TORKLEQR--AELDEMYGQIV-----458	
Qy	664	NKLFKEKEVLYVNSQGEISRLSTKKEVINNNYKLGQEGKYRVYHNOYSTNSFAL	723
Db	459	---QMKQLIROHMAQMEEMKTRHKGEMENALRSYSNI-----TVNEDQIKLMNVAI	507
Qy	724	NKHQR-EDHDKRRH-----LAHKFLCTPAGFEKWNKSGVHGSVKVLTITLRLTITQ	773
Db	508	NELNIKQDINSQEKLEKEELGLILEEKCALQ-----RQ	541
Qy	774	LENNIPSSFPHNWAHRANWIKAVQMCSPREFALALALECAVKPVVMLPIWREFLGH	833
Db	542	LEDIVELSFSEQIQARQTI--AEQESKLNKAHSLSTVEDLKAEIVASERSKELEH	599
Qy	834	TRLHRMTS-----IEREKEKVKKKEEBEETMQATWVYTPPVKHQVWKQGE---E	886
Db	600	KHEAVNYKTKLEMEKKNVLDRAESQEAELERLTQLLFSHEELSKLEDLEIE	659
Qy	887	YRVTYGGWSWISKTHVYRVPKLPNGTNVNYRKSLEGTKNNMDM--ESDKRKSRS	944
Db	660	HRIN-----IEKLDNLGIHYKQOIDLQGNESQKITEQFQKDNLTIK	703
Qy	945	PKKIKIEPDSKEVKGSAAKAD-----QNMEDISKITEK-----DQDYKEL---	989
Db	704	QNQLILEISKLDLQOGLVNSKSEMTLQINELQKEIILRQEEKEKGTLEQVEQLQK	763
Qy	990	-----LSDSDKPCKEEPMVEVDDMKMTESHVNCQSSQSDVVNV	1028
Db	764	TELLEKQMKENDLQEKFAQLEAE--NSILKDEKKTLEDMLKIHTPVSQEE-----	813
Qy	1029	SEGPHLRTSYKTKSSKLDGLLERRIKQFTLEEKORLEK--IKLEGGIKIGKTSNSS	1086
Db	814	-----RLIFLDSIKSKSDSVWEKEI--EIIIEENEDLKQOCIQLNIEEKQRTFSFAE	866
Qy	1087	KNL-----SESPVITKAKEGQSDSMRQE---QSPNANNQDPELQGCQSQSSSV	1134
Db	867	KNFEVNYQLEQEYACLLKVKDDLEDSSKNKQOEYKSKLKALNEE-----911	
Qy	1135	LRMSDPSTHTNKLYPKDRVLDVSTRSPET-----KCPKQNSIENDI	1176
Db	912	LHLQRINPTTVKM--KSSVFDSDKTFVAETLEMGEVVEKDTTELMKLEVTREKLE--L	967
Qy	1177	EKVSDLASRGOEPTKSKTKGNDFTIDDSKLASADDIGTLICKNNKPLIQESDTIVSSS	1236
Db	968	SQRLSDLS-----EQLKQKHGEISFLNEEVKSLQE-----KEQVSLRCRELEIINH	1016
Qy	1237	KSALHSSVPKSTNDRDATPLSRAMDFEGKLGCDSESNSTLNSSTVSIQSDSEDMIVQ	1296
Db	1017	RA-----ENVQSDCTQVSSIL-----DGVVTMTSRGEGSVS	1048
Qy	1297	NSNESISEQFRTREQDVEVLEPLKCELYSGESTGNCE---DRLP--VKGTEANGKPSQ	1351
Db	1049	KVNSFGEEKSWKVEDKVSFE-----NMTVGEESKQELILDHLPSTVTKESSLRATQPSN	1104

```

QY 1352 KLEERPYNKSDQIKLKNWTDKKNENRESKKGOR-----TSTFOI-----NGKD 1398
DQ 1105 DKLQ-----KELNVLKSEQNDLR-LQMEAQRICLSLVYTHVDQVREYMENEXD 1152
QY 1399 NKPKIYLKCEKLEISESVVSGNVPKNNINKIIPENDIKSLTVKESAIRPFINGD-- 1456
DQ 1153 -----KALCSLKEELIFA--QEEKIKELQK-IHQLEQTMKTQET-----GDEG 1193
QY 1457 -----VIMEDFNERNSETSKHLSLSDAEGNYRDSLETLPSTKESDSTOTTPSPASCPS 1512
DQ 1194 KPHLLIGLQKAVSECSYFLQTLCSVLGEY-----YTPALKC-----1232.
QY 1513 NSVNQVEDMEITSEVKVKTSPITSEESNLSNPDIDENGILPKNK-----ENNVGSKRK 1569
DQ 1233 -----EVNAEDKEN--SGDYISENEDPELDQYRYEQDQENMH 1269
QY 1570 TVITEVT-----TMTSTVATESKTVIKVERGDKQTVVSTSTENCASVTITTTTV 1619
DQ 1270 TLLNKVTEYNKLLVLQTLKSLIWGQOTDGMKLEFE-----ENLPKE-----1312
QY 1620 TKLSTPSTGGSDIISVKBSQKTVVTTVTDSLTGTLVTSMTVSKYSTKDKVKLKM 1679
DQ 1313 -----ETEFLSIHQ-----MTNLEDIDVNHKSLSS 1339
QY 1680 FSRPKTRSGTALPSYRKVEVTKSTKSIPLVNDLKLKARKGIREVPYFNYNKAPALD 1739
DQ 1340 LODLEKTKLEEVOQELSISSIQO-----LKETEQ-----NYAE-----1376
QY 1740 IWPYSPRPTFGITWYRLQTVKSLAGVSLMLRLLWASLRWDDMAKVPVPGGSTRTETS 1799
DQ 1377 -----IHCLQ--KRLQAVS-----ESTVPP-----SLP 1397
QY 1800 ETEITTEIKRRDVPYIREPYCIRKILICIGVPETPKETPTORKGLRSSALRPKRP 1859
DQ 1398 VDSVITESDAQRTWYPGS-----CVKKNI--DGIEFSGEFGVKEETNIVKLEKQYE 1450
QY 1860 ETPKQGPVLIETWA--ELEELEWEIRAFARVEKEK--AOAVEQQAQKLEOKPVTIAT 1917
DQ 1451 QLEEEVAKVIVSNAFAOOTSLSRISGKENTASSQAHAQCQEQBYFNEMK-----L 1505
QY 1918 STTSPSTSTISPAQKVMVAPISGSVTTGPKMVLTKKVGSPAIVTFOONKHFOTPAT 1977
DQ 1506 SODQIGQFTFDVVKFEERPLSKELGEHKEILLS-----NSDPHD-----1549
QY 1978 WYKQGSNGVYVQVQKVLGIIPSTGTSTQOFTSFQPRRTATVTRPNTSGSGGTTNSQ 2037
DQ 1550 -----IPESKDCVLT-----SEEMFSKDKTFIVRSIHDELSVSSMDASROL 1592
QY 2038 VITGPIRPGWTVIRTPLOOSTLGKAIITPPVWVOPGAPQVWQTIIRQPVSTAVSAPN 2097
DQ 1593 MLNEEOLED-----MRQELVROYQEHQQA-----1616
QY 2098 TVSSTPGOKSLTSATSTSIQSSASOPPRPOGQVKTWLAQTLTOLTOGHGNOGLTVYIQ 2157
DQ 1617 TORSSIDNENLSEERERVLLEBALKQLSLAGREKLCCCELNSSTQTQNGN-----E 1669
QY 2158 GOGOTTGQ-----LQLIPQGV--TVLPGPQQLMQAAMPNGTVORFLFTPLATTATTAS 2209
DQ 1670 NOGEVEEQTFKEKELDRKPEDVPPELSNERYALOKA---NNRLKILLEVKTAAVEE 1726
QY 2210 T-----TTTTVSTTAAGTGEOQRQSKLSQPMQVHQDKTLP-----PAQSSVGP-----2252
DQ 1727 TIGRVLGLDRSSKSSQSSASLIRSEAEASVKSCVHEHTRVTDIESIPSYSGSDMPRND 1786
QY 2253 ----AKAQPQTAQPSARPQPOQPSAPOE-----VQTOPEVOTQTVSSHVPSEAQ 2301
DQ 1787 INWMSKVTEEGTGLSORLVRSGFAGTEIDPNEELMLNITSRLQAQAVEKLEAISSETSQ 1846
QY 2302 PTHAQSSKQVAAQS-----QPSNVQGSPPRVQSPSOTRIRPSTPSPQSGQ-----2350
DQ 1847 LEHAVTQETLMRESFRQKQKATESLSKQBELRLHEESRAREQLAELSKAEGVIDGY 1906

```

```

RESULT 9
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF.
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; NUMBER OF INVENTION: PORTIONS AND PRODUCTS
; CORRESPONDENCE ADDRESS:
; ADDRESS: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

US-08-714-741-32

Query Match 2.5%; Score 378; DB 4; Length 8991;
Best Local Similarity 16.7%; Pred. No. 6.4e-13;
Matches 592; Conservative 475; Mismatches 1390; Indels 1082; Gaps 132;

1 MYSEEEEDGDA--EETQDSEDD--EEMEDDDDDYPEEMEDDDDDASYCTESSF 55
1545 VASQSAKADDAVAEKAEEDELEEAARQYDEQKSENEKETEAS-----1597

56 RSHSTYSPTGRRKPRVHRPRSPILEEKDIPLEFPKSSSDEM--VPNEHIMNVIAIYEV 113
1598 -----ERQAAATLKYHLESXEFNLNYFQDNHRMNMKKM--I 1630

114 LNFQTVLRLSPFEDFCAALYSQECTLMAEHVLLKAVLREE--DTSNTTGPADLK 172
1631 LTSLASVAILG-----AGLVASPTVVRAREAPVASQSAEKDYDTAK-----R 1674

173 DSVNGLTFIDGMTWPEVL-----RVYCESDKYHHVLPYQEA--EDYPYGPVENKIKVLQ 226
1675 DAENAKKALEAKRAQKAYADYQRIIEKAETHASLEQOEANKDY-----QLKLKK 1727

227 FLVDQFLTNIARELMSEGVQYDDHCRVCHKGLDGLCCETCSAVYHLECVKPLEEVP 286
1728 YLDGRNLSNSSLVKKEMEEAKKKEKPAEFNKI-----RREIVVPNPQELE 1774

287 EDWQCEVCVAHKVPGVTDCAEIOKNPYIRHEPIGYDRSRKRYWFLNRLIIEEDTEN 346
1775 MARRKSEVAKT--KESGLVKRVEEAKKVTARPK---LDAERAKEVVLQAIAM-----1824

347 ENEKKIWTYS-----TKVOLAEI-----IDCLDKDYWEAE-----LCKIL 381
1825 -NKKMILTSLASVAILGAGLVASPTVVRAREAPVASQSAEKDYDTAKRDAENAKKAL 1883

382 EMREIEIHRMDITEDLTNKG-----SFKSFLAA 413
1884 EEAQRAQYADYQRIIEKAETHASLEQOEANKDYQLUKKYLDGRNLSNSSLVKKE 1943

414 NEELIESIRAKKGDIDN-----VKSPEETEKD 440
1944 MEAEAKKKEQAGLNMKKMILTSLASVAILGAGLVTSOPTLVRAEESPVASQSAEKD 2003

441 KNETENDSKDAENREE-----FEDQSLEK-----466
2004 YDAARDAENAKKALEAKRAQKAYADYQRIIEKAEBQQAQSALEQOEANKDYQLKKY 2063

467 SDDKTDDPDEQCKSEPTVEVDKGNVSAN--LGDNTTATSEETSPSGRSPV-----519
2064 LDGRNLSNSSLVKKEMEEAKKKEQKQAEFNKIRREIVVPNPQELEMAARKSEVYKAKES 2123

520 GCLSETPDSSNMAEKV-----ASEL---PODYPEEPNKTCSN-----TSA 559
2124 GLVKRVEE---AEKKVTEARQKLDRAERAKEVVLQPTRVENEVHKLXQKMNKKMILTSL 2179

QY 560 TTTSI-----QPNLENSNSSSELNSQS-----ESAKAADDPNGRGRESHTPVSIQF 606
Db 2180 ASVAILGAGLVTSOPTFVRAEESPVVEKSSLEKKEYEAKAKAD--TAKKDYETAKKKA 2237
QY 607 EIVGDTFSEKSTGSELSESPCAGKAGSSTRII-----638
Db 2238 DAOKKYEDDOKRTE---EKARKEABASOKLIDLVALVQNAKEYREVQNRQSKYKSDAD 2293
QY 639 --TRLRNPDKLSOLKSQOVAAAAHEANKLFKEGKEVLV-----675
Db 2294 YQKLTVEVDSKIEKARKEQ-----QDLONNFNEVRAVAPDPTCVGDXDRMKNKKMILTS 2348
QY 676 -----VNSQGEISRLSTKKEVIMKGNNTNFKLQOEGKYRYVHYNQY--STNSFA 722
Db 2349 LASVAILGAGXVTSOPTXVRAEEAPVVEKSSLEKKYE-EAKAKYDAAKDYDEAKKKA 2407
QY 723 LNKHOHREDHDKRRHLAKHKECLTPAGEFKWNGSVHGSKVLTJSTLRLT-----IT 772
Db 2408 EAOKKYEDDOKKTEKAEK-----AKAASEIAKATEVQKAVLDYIT 2450
QY 773 QLENNIPSSFPHPNWASHRANWIKAVOMCSKPREF-----ALALALECAVXPVVM 823
Db 2451 AIRHNDSGKTSAAEAENKA---KERYCCAGKKFDPQTPFVASLTQMLMKNKKMILT 2507
QY 824 LPIWREFLGHTRLHRWTSIEREEKEKVKKEKQEEETMQ-----OAT 867
Db 2508 SLASVAILGAGLVASSPTVVRAREAPVASQSAEKDYDTAKRDAENAKKALEAKRAQEK 2567
QY 868 WKVYTFPVKHQVWKO-----KGEYRVTYGGWSWISKTHVYRFVFKPLGNTNVN---YR 919
Db 2568 YADYQRIIEKAETHASLEQOE-----ANKDYQLKLKYLDGRNLSNSSLVK 2616
QY 920 KSLGTTKNNMDENNDESK-----938
Db 2617 KEMEEAKKKEQAEFNKIRREIVVPNPQELEMANMKNKKMILTSLASVAILGAGFVASQ 2676
QY 939 ---KCSRSP---KKIKIEPDSEKDEVGSDAAKADQONEMDISKITEKK-DQDVK---987
Db 2677 PTVVRAEESPVASQSAEKDYDAKAKNAKKAVEDAQKALDDAKAAKKYDEQDKTE 2736
QY 988 -----ELLD-----SDSDKPCKEEPMVEVDDMTESHVNCQESSQV 1023
Db 2737 EKAALKAASEEMDKAVAAVQAVLAYQOATDKAAKDAADAKMIDAEKKEEAKTKFTNV 2796
QY 1024 DVNVVSEGFHRTSYKK---KTKSSKLDGLLERRIKIQTLEEKORLE-KIKLGGIKGI 1078
Db 2797 RAMVYVPEPEQLAETKKKSEAKQAPELTKKLEBAKAKLEBAKAKATEAKQKVD-----2850
QY 1079 GKTSTNSSK---NLSESPVITKAKGQSDSMRQESPMANNDQPEDLIQCSQSDSSV 1134
Db 2851 ---AMNKKMILTSLASVAILGAGLVASOPTLVRAEESPVASQSAE-----KDYDAV 2901
QY 1135 LRMSDPSHTTNKLYPK--DRVLDDVSIIRSPETKCPQNSIEN-DIEEKVSDSLASQOEPT- 1191
Db 2902 KK-----SEAAKAYEAKKALEAKVAKKYEDDQKKTEEKALEKEASEIAKATEVQ 2957
QY 1192 -----KSKYTGNDFFIDDSKLASADDIGTLCKNKKPLIQEESDTLVSSSKSALHSSV 1244
Db 2958 QAYLAYQASNAKAAKAMEEQAQRRENEARAKFTTIRTVVYVPEPEQLAETKKKA-----3012
QY 1245 PKSTNDRDATPLSRAMDFEGKIGCDSNSENSTLNSSDTVSITQDSEEDIMVONNESI---1302
Db 3013 -EEAKAKEPKLAKKAAEAKALE-EAEKKAETANPQ-----VDMNKKMILTSLASVAI 3065
QY 1303 -----SEQFTFRQDVEVLEPLKE-----LVSGESTGNCEDRLPVKTEANGKKPS 1349
Db 3066 LGAGEFVASSPTFVRAEEAPVANSQSAEKDYDAVAKKSEAAKKDYATAKKDAQKAYDE 3125
QY 1350 QOKKLEER--PVNKCSDOI-----KLKNTTDKKNENRESE---KKGORT 1389
Db 3126 DQKTEAKAEKKEKASEKIAETKEVQOAYLAYLQASNESQKAEADKKIKEATHAKMRRT 3185
QY 1390 STFQI-----NGKDNPKIYLKGECLKEISESRVVSNGNEPKNV 1429

[illegible]

Db 1802 VPAVITPPAIPUSGRRRTTKSGKSSRSSLGNKSPOLSGNLSQSAASVHLHPQOT- 1860
Qy 2338 IRETPSQLSPGOQSOQVTTTSOPIPIQPHITSLIQIPSOQPO- SQPOVQS- STQTLSSGQ 2395
Db 1861 LHP-----PG-----NIPESGNQLLOPLKPSPPSSDNLYSAF 1892
Qy 2396 TLN-QVSVSPSPRQIQIQOP 2415
Db 1893 TSDGAISVPSLSAPGQGIKOP 1913
RESULT 14
US-09-854-856-62
; Sequence 62, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 2048
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2048)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-854-856-62
Query Match 2.5%; Score 370; DB 4; Length 2048;
Best Local Similarity 18.2%; Pred. No. 2.3e-13;
Matches 419; Conservative 287; Mismatches 746; Indels 846; Gaps 96;
Qy 508 EETSPSEGRSPVCLSETDSSNMAEKKVASELPQDVPEPKNKTCBSSNTSATTTISQPN 567
Db 62 EETVTATATSOVA--QOPPAAPAGQAVAGPAPSTVP-----SSTSKDRPVSPQS 110
Qy 568 LENSNSSELNNSQSESAKAADDPENGERSHTPVSIQEEIVG-----DFTSEKSTG 619
Db 111 LVGSKPEPPPARSGGGGSAKEPQERSQOQDDIELETAKVAGMSNDGRFLKPDIEITGRG 170
Qy 620 ELSSEPCAGKAGSGSTRI-ITRLRNPDLSQLSKSOVAAAAAHEANKLKEGKEVLVNS 678
Db 171 SFT-----VYKGLDTETVEVWCELQDRKLTKSERQ-----FKEEAEMLKGLQ 216
Qy 679 QGEISRLSTKKEYIMKGN--INNYFKLGEGKRVYHNOYSTNSFALNKHQHRD----- 731
Db 217 HPNIVRFYDSTWSTVKGKCIVLVLTMTSGTLTKYLRKFKVMKIKVLRSCWQILKGLQ 276
Qy 732 --HDKRRHLAK-----FCLTPAGEFKWNGSVHSGKVLITSTLRITITQLENNIPSSF 782
Db 277 FLHTRTPPIIHRDLKCDNIFITGPTGSKVIGD-----LGLATLK----- 315
Qy 783 FHNWASHRANWIKAV-----OMCSKPRE-----FALALALECA----- 817
Db 316 -----RASFAKSVIGTFEPMAPMEYEEKYDESVDVYAFMGCMLEMATSEYPPSECON 367
Qy 818 -----VKPV-----VMLPIWRE-----FLGHTRLHRMTSI-- 842
Db 368 AAQIYRRVTSYGVKPSAFDKVAIPEVKEIIEGCIRQNKDERYSIKDLLNHAFFOETGVRV 427
Qy 843 -----EREKEK-----VKKKEKQEEETMQOATWVKYTFPVKHQVWKQGEYR 888

Db 428 ELAEEDGGEKIAIKLWRIEDIKLKGKYDNEA-----IEFSFDERDVPDVAQEMV 481
Qy 889 VTGV---GGMSWISKTHYRVFVKPLPGTNTVNYRKSLGEGTKNNNDENNDSDKRCRSRP 945
Db 482 ESGVCEGDHKTWAKA-----IKDRVSLIKRREORQLVREQEKK----- 522
Qy 946 KIKIEPDEKDEYKSDAAKAGADQNDMDISKITEKKDQDVYKELLDSDKPKCKEPMEV 1005
Db 523 ---KQESSLKQVQEQSSASQ-----TGIKQLPSASTGIP----- 554
Qy 1006 DDDMKTESHYNCQESSQDVVNVSEGHFLRTSYTKKTKSKLKDGLLERRIKQFTLEEKQR 1065
Db 555 ---TASTTSASVSTQVEPEPEADHQLOLQYQOPSISVLSDCGTVD----- 596
Qy 1066 LEKIKLGGIKGIGKSTNSSKNLSESPVITKAKEGQCSQSMRQEQSPNANNDDOPEDLIQ 1125
Db 597 -----SQGSSVFTESRV-----SSQTVYSGSHQAHSTGT 629
Qy 1126 GCSQSDSVLRMSDPSTHTNKLYPKDRVLDDVIRSIRPETKCPKONSIENDI-----EE 1178
Db 630 VPGHIPSTVQAQSQP---HGVTYPPSSVQOGI-----QQTAPPOQTVOYLSQTSSTSEA 680
Qy 1179 KVSOLASRGQEP-----KSKTKGDNDFIDDSKLASADDITGLCKKKNKPLIQEED 1230
Db 681 TTAQPVSPQAQVLPQVVSAGKOSTQG-----VSQVAPAEPAVAV-----AQP--QATQP 727
Qy 1231 TVSSSSKSLHSSVPKSTNDRDATPLSRAMDFEKG----- 1265
Db 728 TTLASSVDSAHSDVASGMSDGNENVPSSGRHGERTTKRHYRKSVRSRHEKTSRPKLR 787
Qy 1266 -LGCDSNSTLENSDT-----VSIQ-----DSSEE-----DMIVONSNESISEQF 1306
Db 788 ILAVSNKGRVVEQCLETHNRKVMVTFKFDLDGDNPEIATIMVNDIFILAIERESFVDQV 847
Qy 1307 RTREQDVEVLEPLKCELVSGESTGNCEDRLPVKGTTEANGKPKSQOKKLEER---PVNKCS 1363
Db 848 R---ETIEKADEMLSEDVSVPEEGD-QGLESLOGKDDYGFSGSKLEGEKQPIPASSMP 903
Qy 1364 DOJLKNWTDKNNENRESEKKGORTSTFOINGKDNKPKLYLKE--CLAEISRSRVSG 1421
Db 904 QQIGIPT-----SSLTQV-----VHSAGRRRIVSPVPSRURES 937
Qy 1422 NVBPKVNNINKIIPENDIKSLTVKESAIRPFINGDIVIMEDFNERNSETSHLSSSDAE 1481
Db 938 KVP-----SEITDTVAASTAQSPGMN---LSHSASSLSLQAASSELRAQWTE 983
Qy 1482 G-----NYRDSLETLPSTK---ESDSTQTTTSPASCPESNSVNOQDEMEIETSEVKVT 1532
Db 984 GPNTAPPNFSGHTGTFVVPFLLSSAGVPTTAAATAPVATSPSPND--ISTV---VI 1037
Qy 1533 SSPITSEESNL-----SNDFIDENGLPIKNENNGESKRKVTITVETMTSTVATESK 1587
Db 1038 QSEVTPTEGIAAGVATSGVTGGGLPI---PPVSESPVLSVSSVSIITPDAVVISITTS 1094
Qy 1588 TVIKKEGDKQTVVSTENCACATSTVTTTTVTKLSTPSTGGSDIISVKEQSKTVVTTT 1647
Db 1095 PSLOQVPTSTSEIVVSSSTALYPSVTSATSA-----SAGGSTATPGKPPA--VVSOQ 1144
Qy 1648 VTOSLTGTLVTMTVSKYSTRDKVKMLKFSRPKTKRSGTALPSYRKFVTKTKSI 1707
Db 1145 AAGS--TTVGATLTSVSTTSPSTASOLS-----QLSSSTSTPTLAETVVVSAH--- 1193
Qy 1708 FVLPNDDLKLRKGGIREVPYFNNAKPAIDWPYSPRPTGITWRYLRQTVKSL--A 1765
Db 1194 -----SLDTSHSS-----TGLAEFSLASAPSSSPGA 1221
Qy 1766 GVSMLRLLWASLRWDDMAAKVPPGGSGSTRTETSETTEITTEIKRRDVGPIGIRFEYCI 1825
Db 1222 GVSSYIS-----QPGG----- 1232
Qy 1826 RTKICIGVDPETPKETP--TPQRKGLRSSALRPKRPTPKOTGPV-----IETWABE-- 1877

Db 1233 ---LHPLVPSVIASTPILPOAAGTSTPLLPQVPSIPPLVQPVANVPAVOQTILHSQPQ 1289
QY 1878 -----LWEIRAFARVE---KEQAQVQQAQKLEQQ-- 1910
Db 1290 PALLPNQPHHCPEVDSQTQKACIDDIKLEKRLSFSEHSSGQAQHASVLSLTV 1349
QY 1911 -KPTV---IATSTTSPT---SSTTSTISPAQV---MVAPISGVTGTGKWLTTTKVGS 1959
Db 1350 IESTVTPGPTAVAPSKLLSTSTCLPPTNLPLGTVALPVTVPVTPG-----QVST 1402
QY 1960 PATVTFQONKHFQTFATWVQOGSNCVVOQKVLGIIPSSGTSTQOQTTSQPRTAT 2019
Db 1403 PVSTT-----TSQVKGPTAPSKPLTKA 1425
QY 2020 VTIRENTSGGTTNSQVITGPOIRPGMTVIRTPLOQ-----STLCK 2062
Db 1426 PVLPGVTELPAGTLPSEQLPFPP--GPSLTOSQPLEDLDAQLRLTSPXITVTSAGVP 1483
QY 2063 AIIRTPVMVOPCA--PQVMTQIIRGQPVSTAVSA-----PNTVSTSPQCKS----- 2107
Db 1484 VSMAAPTALTEAGTQPKGVQVKEGVPVLATSSGAGVFKMGRFQVSVAAQKQEGKNS 1543
QY 2108 -----LTSATSNIOSSASQPRPOQGOVKLTMAQLTOLTQCHGNOGLTVVIOQOG 2160
Db 1544 EDAKSVHPFESTSESSVLSSS-----PESTLVKPE----- 1574
QY 2161 QTTGQLILPQGVTVLPQPGQOLMOAAMPNGTVQRFLEFTPLATTATTATTT----- 2212
Db 1575 -----PNGITI--PGISSDVESA-----KTASEAKSDTGQPTKVGR 1611
QY 2213 TTVSTTAAGTGEQROSK-----LSPQOVHQDKTLPAPASSVGPAPAKAPQ 2258
Db 1612 FOVTTTANKVGRFSVSKTEDKITDTKKEGVPASPPFMDLEQAVLP-----AVIPKKEKE 1666
QY 2259 TAQPSARQPOQPO-----SPAQP----- 2278
Db 1667 LSEPHLNGSPDPEAFLSRDVDDGSGSPHSHPOLSKSLPSQNLSSNSFNSSYMS 1726
QY 2279 -----EVO-----TOPEVQOTQTVSSHP----- 2297
Db 1727 SDNESDIEDLKLRLRLDKHLKLEIQLDSQKHEIESLYTKLGKVPVAVIIPPAAPL 1786
QY 2298 --SEAQPHASSKQVAA---QSOPQ--SNVQGSVPVRQSPQSTRIRPSTPSQLSPGQ 2350
Db 1787 SGRRRRTKSGKSSRSSLGNKSPQLSGNLSGQSAASVLPQQT--LHP--PGNIPBSG 1843
QY 2351 OSOVQTTTSQPIQPHF-----SLQIPSGOGAPQOPQVQSSQTSLSSQOTLNO 2399
Db 1844 QNQL-----LQPLKPSFSSDNLYSFTSDGAISSVPSLSAP---GQCTSTNTV--GATVNS 1894
QY 2400 VSVSSPSRPLQIQOQPQ 2417
Db 1895 -----QAAQAPQ 1901

RESULT 15

US-09-854-856-8
; Sequence 8, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8
; LENGTH: 2229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2229)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-854-856-8

Query Match 2.5%; Score 369; DB 4; Length 2229;
Best Local Similarity 17.9%; Pred.No. 3e-13;
Matches 443; Conservative 306; Mismatches 781; Indels 948; Gaps 111;

QY 472 PDDPEQCKSEBPEVGVGDKGNSVANLGDNTNTATSE-----ETSPSEGRSPVPGCLSTP 526
Db 21 PAPAPKNGSSD--SSVGEK--LGAADAADAVTGRTEEYRRRHMTMDKSRGAAATTTTTE 76
QY 527 -----DSS-----NMAEKVASELPQDVPPEPNKTCSSNTSATTTTSI- 564
Db 77 HRFFRRSVICDSNATALEPLGLPLSLPQSPISPAVPAVQSAPEPHR--EETVTATATSQVA 134
QY 565 -QPNLENSSSSELNSSOSESAKAADDPEN--GERESHTPVYSIOEIVGDFTSEKSTGEL 621
Db 135 QQP-----PAAAAPGEQAVAGPAPSTVPSSTSKDRPVS-QPSLV-----GSK 175
QY 622 SESPCAGKAGSGSTRIITRLRNPDLSKLSQKSSQVAAAAHEANKLFKEGKLYLVNSQGE 681
Db 176 EEPFARSQSGGS-----AKEPQERSQ-----QQDD 203
QY 682 ISRLSTKKEVIMKGNINNYFKLQBP---GKRVYHNYSTNSFA-----LN 724
Db 204 IEELETK--AVGMSNDGRFLKFDIEIGRGSFKTVYKGLDTETTVEVAMCELODKRLTKSE 261
QY 725 KHOHREDHDKRRHLAH-----KFCLTTPAGEFKWNGS---VHGSKVLTI 764
Db 262 RQRFKEEAEMKLGQHPNIVRFYDSWESTVVGKKCIVLTELMTSGTLTKYLKRFKVMKI 321
QY 765 STLRLTITQLENNIPSSFFHPNW-----ASHRANW 794
Db 322 KVLRSWCQILKGL--QFLHTRTPIIHRDLKCDNIFITGPTGKVGIGDLGLATLKRSF 379
QY 795 IKAV-----QMSKPRE-----FALALAILECA-----V 818
Db 380 AKSVIGTPEFMAPEYEEKYDESVDVYAFGCMCMLEMATSEYPSYEQNAQIYRVTSV 439
QY 819 KPV-----VMLPILWRE-----FLGHTLRHMTSI-----EREKEK-- 849
Db 440 KPASFQKVAIPVKIIEGCIQONKDERYSIKDLLNHAFQOETGVRVLEAEDDGEKIA 499
QY 850 -----VKKKKEKQEEETMQOATWVYTFPVKQVWKQKGEYRVYGY---GGWSW 897
Db 500 IKLWLRIEDIKKLGKYKDNEA-----IEFSFDERVDPEDVQAQEMVESGVCEGDHKT 553
QY 898 ISKTHVYRFVPLPGNTWVNYRKSLGCTKNNDENM--DESDKRRKCSRSPKKIKIEPDSE- 955
Db 554 MAK-----IKDRVSLIKRKRQORQVIREEQEKKKQOESLSKQOQVQSSAS 599
QY 956 ----KDEVKGSAAKAGADQNMEDISKITEKKDQDKELLDSKDPCKPEEPMEVDDDD--- 1008
Db 600 QTGIKQLPSASTGIPTASTTSASVSTQVEPEPEA---DQHQOQLQYQPSISVLSDGTV 655
QY 1009 -----MKTESHVNCQES-----SOVDVVNVSEGHRTSYKK 1040
Db 656 DSGQGSQSVFTESRVSSQQTVSYSQHEQAHSHTGTVPGHIPSTVQAQSPGHVYPPSSVAQ 715
QY 1041 -----KTKSKLDGLLERRIKOFTLEEKORLEKIKLEGGIK-----GICKTSN 1084
Db 716 GQSQGQPSSSSLTG-----VSSQPTOHPOQOQGIQTAPQOQTVQVSLSQTSIS 765
QY 1085 SKKNLSES-----PVITKAKGQGSQSMRQEQSPNANNNDPEDLIQ--CQSQSDS 1133
Db 766 SEATTAQVPSQPAQVPLPOVSAGKQG--FPRLPPLQPYGDSNIAPSSNVASVCIH--ST 821

QY	1134	VLKMSDPST-----TNKLYPKDRVLDDVSIKSPETKCPKQNSIENDIE	1177
Db	822	VLXPPMPTVELATPGYFPTVQVYVESNLVLPMGVGGVQVQVQPG-----	867
QY	1178	EKVSDLASRGOEPTKS-----KTGNDFFIDSKLASADDIGTLICKNKKPLIOESD	1230
Db	868	-----GSLAQAPTSSQOAVLESTOG-----VSQVAPAEPAV-----AQP--QATQP	908
QY	1231	TIYSSKSALHSSVPKSTNDRDATPLSRAMDFEKG-----	1265
Db	909	TTLASSVDSAHSDVAGMSDGNENVPSSGRHEGRTTKRHYKRSVRSRSHRSTSPKLR	968
QY	1266	LCGDESNSSTLENSDT-----VSIO-----DSSEE-----DMIVONSNESTSEOF	1306
Db	969	ILNVSNKGRDVRVEQLETHNRKMVTTFKFDGDNPEEIAITIMVNNDFILAIERESFVQV	1028
QY	1307	RTREQOEVLEPLKCELVSGESTGNCEDRLPVKTEANGKKPKQKLEER-----PVNKCS	1363
Db	1029	R---EIIERADEMLSEDVSEPEGD--QGLESLQGDYGFSGSQKLEGEFKQIPIASSMP	1084
QY	1364	DQIKLNTTDDKNNENRESEKKGQRTSTFOJGDKNKPXYLKGE--CLKEISESRVVS	1421
Db	1085	QOIGIPT-----SSLTQV-----VHSAGRRFIVSPVESRLRES	1118
QY	1422	NVEPKYNNIKIIPENDIKSLTVKESAIRPFINGDIVMEDFNERNSETKSHLLSSDAE	1481
Db	1119	KVEFP-----SEITDVAASTAQSPGMN---LSHSASSLSLQQAFLSELRRQOMTE	1164
QY	1482	G-----NYRDSLETLPSTK---ESDSTOTTPSASPESNSVNOVEDMEIETSEVKKVT	1532
Db	1165	GPNTAPNPSHTGTFPVVPPFLSSLAGVPTTAATAPVATSPSPND--ISTS-----VI	1218
QY	1533	SSPITSEESNL-----SNDFIDENGLPINKNENNVNGESKRKTIVITEVTMTSTVATESK	1587
Db	1219	QSEVTVPTBEGIAGVATSGVNTSGGLPI---PPVSESPVLSWSSSITIPAVVSISTS	1275
QY	1588	TVLKEGDKQTVVSVSTENCASKSTVTTTIVTKLSTPSGGSVDDIISVKEQSKTVVTTT	1647
Db	1276	PSLQVPTSTSEIVSVSTALYPSVTVSATSA-----SAGGSTATPGKPPA--VVSQ	1325
QY	1648	VTDLSLTGTTGLVTSMTVSKEYSTRDKVKLMLKFSRPKTRSGTALPSYKRFVTKSKKSI	1707
Db	1326	AAGS--TTVGATLTVSTTTSFPTAQSLI-----QLSSSTSTPLAETVVVSAH---	1374
QY	1708	FVLPNDLKLARKGGIREVPYNYNAKPAIDLTWYPSPRPTFGIYWRVLOTVKSL--A	1765
Db	1375	-----SLDKTSHSS---TTGLAFSLAPSSSSSPGA	1402
QY	1766	GVSLMLRLLWASLRWDDMAKVPPGGSTRTETSETTEITTEIIKRDVCPYIREYCI	1825
Db	1403	GVSSYIS-----QPGG-----	1413
QY	1826	RKIICPIGVPETKPTP--TPQRKGLRSSALRPKRPETPKQTGPV-----IIETWAE--	1877
Db	1414	---LHPLVIPSVIASTPILPQAAGPTSTPLLPQVSPILPVPVANVPAVQOQLIHSQPQ	1470
QY	1878	-----ELE-----LWETRAFAERVE---KEKAQAVEQOAKRLBSQ--	1910
Db	1471	PALLPNQPHTHCPEVSDSQKAPAGDDIKTLEKRLSFLSEHSSSGAOSHVSLETSLV	1530
QY	1911	-KPTV---IATSTSTPT---SSSTSTISPAKV---MVAPISGSVTTGKWLTKVGS	1959
Db	1531	IESTVTPGPTTAVAPSKLLTSTSTCLPTNLPGLTVALPVPVPTPG-----OVST	1583
QY	1960	PARVTFPQKNFHOTFATVWKQGSNGVVVQOKVLGIIPSSGTSTSQOTFTTSFQPRAT	2019
Db	1584	PVSTT-----TSGVKPGTAPSKPPLTKA	1606
QY	2020	VTIKPTVSGSGTTSNSQVITGQIIRGMTVIKRPLQO-----STLKG	2062
Db	1607	PVLPGVTELPAGTLPSEQLPPFP--GPSLTSQOQPLEDDAQLRRLTSPSEXIVTISAVGP	1664

Qy	2063	AIITPVMVOPGA--POQVMTQIIRQOPVSTAVSA-----PNTVSSTPGOKS-----	2107
Db	1665	VSMAAPTAITAEACTPQOKGVSYQVKEGVLATSSGAGVFKMRGFQVSAADGAQKEGKNKS	1724
Qy	2108	-----LTSATSTSNTOSSASOPPRPQOQOVKLTMAQLTQLTQGHGGNOGLTVVIOGQG	2160
Db	1725	EDAKSVHFESSTSESSVLSSS--PESTLVKPE-----	1755
Qy	2161	QTTQOLQOLIPOGVTVLPGCQQLMQAAMPNGTVQORFLTPTLATTAATASTTT-----	2212
Db	1756	-----PNGITI-EGISSDVPESAH-----KTTASEAKSDTGQPTKVGR	1792
Qy	2213	TTVSTTAAGTGEORQSK-----LSPQMQVHDKTLPPAQSSVGPAPAKAQP	2258
Db	1793	FOVTTTANKVGRESVSKTEDKITDTKEGVPASPPFMDLEQAVLP-----AVTPKKEPKE	1847
Qy	2259	TAQBSARPQOTQPO-----SPAQP-----	2278
Db	1848	LSEPSHLNGPSSDPEAAFLSRDVGSGSPHQLSSKSLPSONLSQSLNSFNSSYMS	1907
Qy	2279	-----EVO-----TOPEVOTQTVSSHVP-----	2297
Db	1908	SDNESDIEDEDLKLELRLRDKHLKEIQDLOSQKHEIESLYTKLCKVPVAVIIPPAAPL	1967
Qy	2298	--SBAQPTHAKSSKPQVAA--OSQPO--SNVOGQSPVRVQSPQSTRIIPSTPSQLSPQG	2350
Db	1968	SGRRRPTKSGKSSRSRSSILGNKNSPOLSNLSCQSAASVLHPQOT-LHP--PGNIPESG	2024
Qy	2351	QSOVQTTTSPPIQPIQHT-----SLOJTPSQGQPOQPOVOSSTQTLSSGOTLQAO	2399
Db	2025	QNQLU-----LOPLKPSPSDNLISAFSTSDGAISVPSLSAP---CQGHSTSTNTV--GATVMS	2075
Qy	2400	VSVSSSRPQLQIQQPQ 2417	
Db	2076	-----QAQAQAP 2082	

Search completed: September 24, 2003, 01:18:58
Job time : 53.9984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 00:30:12 ; Search time 69.5063 Seconds
(without alignments)
10792.671 Million cell updates/sec

Title: us-09-698-295-1
Perfect score: 14971
Sequence: 1 MYSEEEEDGDAETQDSE.....KLKGFASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	14170	94.6	2781	4 Q9UIG2	Q9uiq2 homo sapien
2	3798	25.4	796	11 O8CFX5	O8cfx5 mus musculu
3	3445	23.0	803	11 O8K092	O8k092 mus musculu
4	2781.5	18.6	2649	5 Q9W0T2	Q9w0t2 drosophila
5	2775.5	18.5	2669	5 Q9W0T1	Q9w0t1 drosophila
6	2771.5	18.5	2669	5 Q95VB8	Q95vb8 drosophila
7	2705.5	18.1	645	11 Q8VDN7	Q8vdn7 mus musculu
8	1368.5	9.1	412	4 Q9H5E0	Q9h5e0 homo sapien
9	854.5	5.7	1711	5 Q45A09	Q45a09 caenorhabdi
10	848.5	5.7	1713	5 Q95XU8	Q95xu8 caenorhabdi
11	692.5	4.6	1022	5 Q960Y3	Q960y3 drosophila
12	641	4.3	7210	5 Q9V7G8	Q9v7g8 drosophila
13	641	4.3	9270	5 Q8MLD9	Q8mld9 drosophila
14	621.5	4.2	2768	5 Q9VC00	Q9vc00 drosophila
15	613.5	4.1	5476	5 Q9NJ17	Q9nj17 drosophila
16	613.5	4.1	5533	5 Q9VPL2	Q9vpl2 drosophila

17	613.5	4.1	5554	5 Q9NHN1	Q9nhn1 drosophila
18	613.5	4.1	5560	5 Q9VPL1	Q9vpl1 drosophila
19	608.5	4.1	5533	5 Q9U6C3	Q9u6c3 drosophila
20	593	4.0	17352	5 Q95YM2	Q95ym2 procambarus
21	574	3.8	18074	5 Q9I7U4	Q9i7u4 drosophila
22	571.5	3.8	16215	5 Q9NFS3	Q9nfs3 drosophila
23	562.5	3.8	3257	5 Q9V736	Q9v736 drosophila
24	546.5	3.7	5327	5 Q76891	Q76891 drosophila
25	535	3.6	5412	5 Q9W596	Q9w596 drosophila
26	535	3.6	16274	5 Q8IR22	Q8lr22 drosophila
27	520.5	3.5	10578	5 Q8ISF5	Q8isf5 caenorhabdi
28	518.5	3.5	2081	10 Q9LH98	Q9lh98 arabidopsis
29	517.5	3.5	18519	5 Q8ISF6	Q8isf6 caenorhabdi
30	517.5	3.5	18534	5 Q8ISF7	Q8isf7 caenorhabdi
31	517	3.5	99	11 Q8C473	Q8c473 mus musculu
32	508	3.4	2910	10 Q9FND5	Q9fnd5 arabidopsis
33	504.5	3.4	2465	5 Q8I071	Q8iq71 drosophila
34	503.5	3.4	4498	5 Q9W223	Q9w223 drosophila
35	501	3.3	2310	16 Q8CM07	Q8cmu7 staphylococ
36	496	3.3	6994	5 Q17343	Q17343 caenorhabdi
37	490	3.3	6994	5 Q17490	Q17490 caenorhabdi
38	489	3.3	2402	2 Q9AER7	Q9aer7 staphylococ
39	487	3.3	3443	11 Q8JZM8	Q8jzm8 mus musculu
40	475.5	3.2	2759	5 Q9VID9	Q9vid9 drosophila
41	474	3.2	3358	5 Q8INF9	Q8inf9 drosophila
42	474	3.2	3726	5 Q9VFL1	Q9vfl1 drosophila
43	460.5	3.1	3484	5 P91257	P91257 caenorhabdi
44	456.5	3.0	3261	4 Q9Y556	Q9y556 homo sapien
45	456.5	3.0	3664	4 Q96T58	Q96t58 homo sapien

ALIGNMENTS

RESULT 1				
Q9UIG2	PRELIMINARY;	PRT;	2781 AA.	
ID	Q9UIG2			
AC	Q9UIG2;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Bromodomain PHD finger transcription factor.			
GN	BPTF.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20130111; PubMed=10662542;			
RA	Jones W.H., Hamada N., Shumane M.,			
RT	"Identification and characterization BPTF, a novel bromodomain			
RT	transcription factor."			
RL	Genomics 63:35-39(2000)			
DR	EMBL; AB032251; BAA89208.1; -			
DR	HSSP; Q92831; I891.			
DR	InterPro; IPR001487; Bromodomain.			
DR	InterPro; IPR004022; DDT_dom.			
DR	InterPro; IPR006209; EGF_like.			
DR	InterPro; IPR001965; Znf_PHD.			
DR	Pfam; PF00439; bromodomain; 1.			
DR	Pfam; PF02791; DDT; 1.			
DR	Pfam; PF00628; PHD; 2.			
DR	SMART; SM00297; BROMO; 1.			
DR	SMART; SM00571; DDT; 1.			
DR	SMART; SM00249; PHD; 2.			
DR	PROSITE; PS00633; BROMODOMAIN_1; 1.			
DR	PROSITE; PS00014; BROMODOMAIN_2; 1.			
DR	PROSITE; PS01186; EGF_2; 1.			
DR	PROSITE; PS00016; ZF_PHD_2; 2.			
SQ	SEQUENCE 2781 AA; 311210 MW; 7B3BE0C1C4C3CFBE CRC64;			

Query Match 94.6% Score 14170; DB 4; Length 2781;

Best Local Similarity 95.18; Pred. No. 0; Matches 13; Indels 126; Gaps 1;
Matches 2766; Conservative 2; Mismatches

Qy	1	MVSEEEEDGDAEETQDSEDEDEMEEDDDDDSDYPEEMEDDDDDASYCTESSFRSHST	60
Db	1	MVSEEEEDGDAEETQDSEDEDEMEEDDDDDSDYPEEMEDDDDDASYCTESSFRSHST	60
Qy	61	YSSTPGRKRPRVHRPRSPILIEEKDIPPLEPFPKSSSEDLMPVNEHIMNVIAIYEVLRNEGTV	120
Db	61	YSSTPGRKRPRVHRPRSPILIEEKDIPPLEPFPKSSSEDLMPVNEHIMNVIAIYEVLRNEGTV	120
Qy	121	LRLSPFREDECAALVSOEQCTLAEMHVLLKAVLREEDTSNTFGPADLKDSVNSTLY	180
Db	121	LRLSPFREDECAALVSOEQCTLAEMHVLLKAVLREEDTSNTFGPADLKDSVNSTLY	180
Qy	181	FDGMTPEVLRVYCESDKYHHVLPYQAEADYPYGPVENKIKVQLFVQGLFNTNARE	240
Db	181	FDGMTPEVLRVYCESDKYHHVLPYQAEADYPYGPVENKIKVQLFVQGLFNTNARE	240
Qy	241	ELMSGVIOYDDHCRVCHKLGLLCCETCSAVYHLECVKPPLEVPDEDMOCEVCVAHKV	300
Db	241	ELMSGVIOYDDHCRVCHKLGLLCCETCSAVYHLECVKPPLEVPDEDMOCEVCVAHKV	300
Qy	301	PGVTCVAEIOKNKPYIRHEPIGYDRSRKRYWFLNRLLIIEEDTENENKKIWYYSTKVQ	360
Db	301	PGVTCVAEIOKNKPYIRHEPIGYDRSRKRYWFLNRLLIIEEDTENENKKIWYYSTKVQ	360
Qy	361	LAELIDCLDKOYWEALCKILBEMREETHRHMDITEDLTNKARGSKSFLAANEEILLES	420
Db	361	LAELIDCLDKOYWEALCKILBEMREETHRHMDITEDLTNKARGSKSFLAANEEILLES	420
Qy	421	IRAKKGIDINVASPETEKDKNETENDSKDAENREEFEDQSLKDSDDKTDDDPDEQK	480
Db	421	IRAKKGIDINVASPETEKDKNETENDSKDAENREEFEDQSLKDSDDKTDDDPDEQK	480
Qy	481	SEPTVEVGDKGNSVANLGDNTTNATSETPSSEGRSPVGLSETPDSNMAEKVASEL	540
Db	481	SEPTVEVGDKGNSVANLGDNTTNATSETPSSEGRSPVGLSETPDSNMAEKVASEL	540
Qy	541	PQDVPPEPNKTCSSNTSATTTTSIQPNLENSNSSSELSSQSESAAKADDPENGRESHT	600
Db	541	PQDVPPEPNKTCSSNTSATTTTSIQPNLENSNSSSELSSQSESAAKADDPENGRESHT	600
Qy	601	PVSIQEEIVGDFTEKSTGELSPGAGKGAGSGSTRIITRLRNPDSKLSQSQVAAAA	660
Db	601	PVSIQEEIVGDFTEKSTGELSPGAGKGAGSGSTRIITRLRNPDSKLSQSQVAAAA	660
Qy	661	HEANKLFKEGKVLVYNVNSOGEISRLSTKKEVINMGNNYFKLQGEKGYRVYHNOYSTNS	720
Db	661	HEANKLFKEGKVLVYNVNSOGEISRLSTKKEVINMGNNYFKLQGEKGYRVYHNOYSTNS	720
Qy	721	FALNKHQHRDHDKRRHLAHKCLTPAGEFKWNGSVHGSKVLTISTLRLTIITQLENNIPS	780
Db	721	FALNKHQHRDHDKRRHLAHKCLTPAGEFKWNGSVHGSKVLTISTLRLTIITQLENNIPS	780
Qy	781	SFFHPNASHRANWIKAVQCMKSPREFALALALECAVPVVMPLTWREFLGHTRLHRMT	840
Db	781	SFFHPNASHRANWIKAVQCMKSPREFALALALECAVPVVMPLTWREFLGHTRLHRMT	840
Qy	841	SIEREEKEKVKKEKQOBEETMQOATWVKYTFPVKHQVWKQKGEYRVTYGGWSWISK	900
Db	841	SIEREEKEKVKKEKQOBEETMQOATWVKYTFPVKHQVWKQKGEYRVTYGGWSWISK	900
Qy	901	THVYRFYKPLPGNTNVNYRKSLEGTKNMNMENDESDDRKCSRSPKKIKIEPDSEKDEVK	960
Db	901	THVYRFYKPLPGNTNVNYRKSLEGTKNMNMENDESDDRKCSRSPKKIKIEPDSEKDEVK	960
Qy	961	GSDAAKADQONEMDISKITEKQDQVKELLDSDSKPCKEPEMEVDDDMKTESHVNCQES	1020
Db	961	GSDAAKADQONEMDISKITEKQDQVKELLDSDSKPCKEPEMEVDDDMKTESHVNCQES	1020
Qy	1021	SOVDVNVSGFHLRTSYKKTKSSKLDGLLERRIKOFTLEEKQRLKLEGGIKGK	1080
Db	1021	SOVDVNVSGFHLRTSYKKTKSSKLDGLLERRIKOFTLEEKQRLKLEGGIKGK	1080

Db	895	SQVDVVNVSEGPHLRTSYKKTKSSKLDGLLERRIKOFTLEEKQRLKLEGGIKGK	954
Qy	1081	TSNSSKNLSESPVITKAKEGQSDSMROEQSPNANNQPEDLIOCCSOSDSVLRMSDP	1140
Db	955	TSNSSKNLSESPVITKAKEGQSDSMROEQSPNANNQPEDLIOCCSOSDSVLRMSDP	1014
Qy	1141	SHTTNKLYPKDRVLDVDSIRSPETKCPKONSTENDIEEKVSDLASRGQPTKSKTKGNDF	1200
Db	1015	SHTTNKLYPKDRVLDVDSIRSPETKCPKONSTENDIEEKVSDLASRGQPTKSKTKGNDF	1074
Qy	1201	FIDDSKLASADDIGTLICKNKPLIOESDTIVSSSKSALHSSVPKSTNDROATPLSRAM	1260
Db	1075	FIDDSKLASADDIGTLICKNKPLIOESDTIVSSSKSALHSSVPKSTNDROATPLSRAM	1134
Qy	1261	DFEGKLGCDSESNSTLENSSDTVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK	1320
Db	1135	DFEGKLGCDSESNSTLENSSDTVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK	1194
Qy	1321	CELVSGETGNCEDRLPVKGTTEANGKKPSQOKLEERPYNKCSDOIKLKNTTDDKKNNENR	1380
Db	1195	CELVSGETGNCEDRLPVKGTTEANGKKPSQOKLEERPYNKCSDOIKLKNTTDDKKNNENR	1254
Qy	1381	ESEKKGORTSTFOINGKDNKPKIYLKGECLKEISESRVVSNGVEPKVNNINKIIPENDIK	1440
Db	1255	ESEKKGORTSTFOINGKDNKPKIYLKGECLKEISESRVVSNGVEPKVNNINKIIPENDIK	1314
Qy	1441	SLTVKESAIRPFFINGDVIMEDFERNSESTKSHLLSSDAEGNROSLETPSKESDST	1500
Db	1315	SLTVKESAIRPFFINGDVIMEDFERNSESTKSHLLSSDAEGNROSLETPSKESDST	1374
Qy	1501	QTTTPSASCPSNSVNOVEDMETSESVKVTSSPITSEESNLNSDFIDENGLPINKNE	1560
Db	1375	QTTTPSASCPSNSVNOVEDMETSESVKVTSSPITSEESNLNSDFIDENGLPINKNE	1434
Qy	1561	NVNGESKRKTIVTEVTMTSTVATESKTIVKEKGQTVVVSSTENCAKSTVTTTTT	1620
Db	1435	NVNGESKRKTIVTEVTMTSTVATESKTIVKEKGQTVVVSSTENCAKSTVTTTTT	1494
Qy	1621	KLSTPSTGGSDVLIISVKEQSKTVTTTDSLTITTTGTLVTSMTVSKYESTRDKVKLMKF	1680
Db	1495	KLSTPSTGGSDVLIISVKEQSKTVTTTDSLTITTTGTLVTSMTVSKYESTRDKVKLMKF	1554
Qy	1681	SRPKKTRSGTALPSYKRVFTKSTKKSIFVLPNDLKKLARKGGIREVPYFYNNAKAPALDI	1740
Db	1555	SRPKKTRSGTALPSYKRVFTKSTKKSIFVLPNDLKKLARKGGIREVPYFYNNAKAPALDI	1614
Qy	1741	WPYPSRPTFGITWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGSGSTRTETSE	1800
Db	1615	WPYPSRPTFGITWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGSGSTRTETSE	1674
Qy	1801	TEITITTELIIRKRDVGPYIGIRFEYCIKIIICPIGVPETPKETPTQKRLRSALLRKRPE	1860
Db	1675	TEITITTELIIRKRDVGPYIGIRFEYCIKIIICPIGVPETPKETPTQKRLRSALLRKRPE	1734
Qy	1861	TPKQTPGVIIETWAAEELWEIRAFARVEKEKAQAVEQOAKRLEQKQPTVIATSTT	1920
Db	1735	TPKQTPGVIIETWAAEELWEIRAFARVEKEKAQAVEQOAKRLEQKQPTVIATSTT	1794
Qy	1921	SPTSSTTSTISPAQKVMVAPISGSVTTCTKMVLTTTKVGSPTATVTFQONKNEHQTFATWVK	1980
Db	1795	SPTSSTTSTISPAQKVMVAPISGSVTTCTKMVLTTTKVGSPTATVTFQONKNEHQTFATWVK	1854
Qy	1981	QGOSSNGVVVOQKVLGIIIPSSSTGTSSQOTFTSFQPRATVITIRPNTSGSGGTTNSQVIT	2040
Db	1855	QGOSSNGVVVOQKVLGIIIPSSSTGTSSQOTFTSFQPRATVITIRPNTSGSGGTTNSQVIT	1914
Qy	2041	GPOIRPGMTVIRTPLOQSTLGAIIIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNVS	2100
Db	1915	GPOIRPGMTVIRTPLOQSTLGAIIIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNVS	1974
Qy	2101	STPGOKSLTSATSTNIOSSASOPPRQOGQVKLTMAQILTQLTQGHGNGOGLTVVIOGOG	2160
Db	1975	STPGOKSLTSATSTNIOSSASOPPRQOGQVKLTMAQILTQLTQGHGNGOGLTVVIOGOG	2034

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032289; AAH32289.1; -.
KW Hypothetical protein.
FT NON_TER
SQ
Query Match 23.0%; Score 3445; DB 11; Length 803;
Best Local Similarity 84.6%; Pred. No. 6,7e-139;
Matches 680; Conservative 44; Mismatches 68; Indels 12; Gaps 3;
QY 1504 TPSASCPSNSVNOVEDMETSETSEVKVTSPITSEESNLNDIFIDENGLPINKNENVN 1563
DB 5 TPRAPCDPNSLSQVDEMETESPEVKRVPSPVGTGGSLSGKGFMDNGLPSSKDENVN 64
QY 1564 GESKRKVTITVETMTSTVATESKTVIKVEKDKQTQVVSSTENCAKSTVTTTTVTVKLS 1623
DB 65 GESQKVTITVETMTSTVATESKTVIKVAKGDKQTQVVSSTENCAKSTVTTTTVTVKLS 124
QY 1624 TPSGGGVDDIISVKEQKTVTTTVDLSLTITGGTGLTVMSTVSKYESTRDKVKLMKFSRP 1683
DB 125 TSPDPTGVTISVKEQKTVTTTVDLSLTITAGTGLTVMSTVSKYESTRDKVKLMKFSRP 184
QY 1684 KTRSGTALPSYKRVFTKSTKSFVLPNDLKLARKGGIREVPYFYNNAKPAALDIWPY 1743
DB 185 KTRSGTALPSYKRVFTKSTKSFVLPNDLKLARKGGIREVPYFYNNAKPAALDIWPY 244
QY 1744 PSPRFTGITWYRLQTVKSLAGVSLMLRLWLWASLRWDDMAKVPVPGGGSTRTTSETI 1803
DB 245 PSPRFTGITWYRLQTVKSLAGVSLMLRLWLWASLRWDDMAKVPVPGGGSTRTTSETI 304
QY 1804 TTTEILKRDVGPYIREYKIRKICIPGPETPKETPTQKGLRSSALRPKRPETPK 1863
DB 305 TTTEILKRDVGPYIREYKIRKICIPGPETPKETPTQKGLRSSALRPKRPETPK 364
QY 1864 QTGPVITITWAAEELEWEIRAFARVEKEKAQAVEQAKRLEQQKPTVIATSTTSPT 1923
DB 365 QTGPVITITWAAEELEWEIRAFARVEKEKAQAVEQAKRLEQQKPAVIAASTTSPA 424
QY 1924 SSTTSTISPAQKVMVAPISGVTGCKMVLTKVGSPTATVTFQONKNEHQTFATVVKOG 1983
DB 425 NNTSSTVSPAQKVMVAPISGVTGCKMVLTKVGSPTATVTFQONKNEHQTFATVVKOG 484
QY 1984 SNSGVVQVQKVLGIIPSTGTSQOTFTSFQRTATVIRNPTSGSGGTTNSQVITGPQ 2043
DB 485 SNSGMVQVQKVLGIIPSTGTSQOTFTSFQRTATVIRNPTSGSAGTTTSQVITGPQ 544
QY 2044 IRPGMTVIRTPLOQSLTKAIIIRTPVMVQGPQVMTQIRGQPVSTAVSAPNTVSTP 2103
DB 545 IRPGMTVIRTPLOQPALKAIIIRTPVMVQGPQVMTQIRGQPVSTAVSAPNTVSTP 604
QY 2104 GKSLTSATSTNSIOSSASQPPRQGGVQKLTMAQLTOLTOGHGNGQGLTVVIOGGQGT 2163
DB 605 VQKGLTPGAAGLPLOPSAPHSPPRQGGVQKLTMAQLTOLTOGHGNGQGLTVVIOGGQGT 664
QY 2164 GOLQILPQGVTVLPQGGQLMQAAMPNGTVQRFLETPLATATTAS----TTTTVTSTA 2219
DB 665 GOLQILPQGVTVLPQGGQLMQAAMPNGTVQRFLETPLATTSATASSSSNSSTTTNATA 724
QY 2220 AGTGQRQKLSFQMQVHDKTLPLPAQSSVGPQAKAQTAQPSARPQOTQPSQPAQPE 2279
DB 725 AGSGEQKQKILPQTVQVATTTLAPTQSSSVPAEQPQAPQAPQPP--QPQPPAQPE 782
QY 2280 VQTQPEVQVQTTVSSVHPSEAQPT 2303

Db 783 VQTQPEVQVQTTVSSVHPSETOPS 800
RESULT 4
Q9W0T2 ID Q9W0T2 PRELIMINARY; PRT; 2649 AA.
AC Q9W0T2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG32246 protein.
DE CG10894.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dugan-Rocha S., Dunn P.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
[2]
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Dou P.L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Matti B., Moshrefi A.,
RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragay V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,


```
QY 1483 -----NYRSL-----ETLPSTK-----ESDSTQTTTPSASCPESYNNOVEDM- 1521
Db 1060 ARAHAVPFSLLTECILVQOECVNTNKKIQEYNASSGCWTFDSTQDSKDIYIESMD 1119
QY 1522 -----EITSVKVKTSPITSEESNL-----SNDFID-----ENGLP 1555
Db 1120 VCSNVEIESTEDSVTGLNSGNAEDVMTPCWRRKRKNQSKSKSYIGTKVDLQDLDKDP 1179
QY 1556 INKNENVNESKRKTVTEVTMTSTVATESKTVIKVEKGDQTVVSTENCASVTWTT 1615
Db 1180 LNK-----QNRFFPIT-----ARPVKREC-----VKYERETFNGERNV----- 1214
QY 1616 TTTVTKLSTPSTGGSV-----DIISVKEQSKTVTTTVDTSLTGTGGTLVTSMTVKEYST 1671
Db 1215 -----YSTSPRGVYLLNDAKLYEQ-----VKT 1240
QY 1672 RDKVKLMKFSRPPKTRSGTALPSYKFKVTKSTKSIIFVLPNDLKLARKGGIREVPYFN 1731
Db 1241 EDKSTITK--KPSYSR-----YPLISNFTLHKKRSLLVLPREFELLKLARLGGKSSINGFH 1294
QY 1732 YNAKPAIDINWPPSRPTFGITWYRLQTVKSLAGVSLMLRLWASLRWDDMAKVPPGG 1791
Db 1295 HAAKN--TIWQYQSRPLFTCWSTYSNATSLSLALQKLWLSCLURWDDMIAK--PPST 1352
QY 1792 GSTRTETSETITTEIIRKRDVGYPGIRFEYCIRKIIICPIGVPTPKETPTPQKGLRS 1851
Db 1353 DGKHQVTTDTETVITLLELLKLRHSGRYGKTSYLRKXVVIPLMKPTVREV--TSIRSLR- 1410
QY 1852 SALRPRKPTPKQTPGVITETWABEELEWEIRAEERVEKEKAQAVEQO--AKRRLAQ 1910
Db 1411 ----KRAESQPTPEQITEWVDEDKLELWEIKFPMGEQEKARLSAVTRSVASROLE-- 1465
QY 1911 KPTVTATSTSTSTSTISPAQKVMYAP--ISGSVTTGTGMVLTKVGSPTATVFOQNK 1969
Db 1466 ----ASGNSGNTSTNGALGVAGRVQLAPKLSEDK--EKMEQOLK--QRAVHQOR 1515
QY 1970 NFHQTFATVWVGQSGNSGVVQOQVLG-----IIPSTSTGTS-----QQTFTSFQ-----PRTA 2018
Db 1516 ----LVATGEITRSVTPVKGVIGSRRVIVKPNPDGTTTIIQCAVTVQVSRGTGGANTA 1567
QY 2019 TWTIRNTSGSGGTSTNS-----QVITGPO-----IPGMTVTRTP-----LOO 2057
Db 1568 AAAAPTGVGGSTQSNPSTSTPHKVQIIRGPDGKVSVRGLNPGQOLVOMPDGKLHVLT 1627
QY 2058 STLGLKALIRTPVMVOPGAPOQVMTQIIRGQPVSTAVSAPNTVSTPGQKSLTSTATSTNI 2117
Db 1628 TTSSNSAGOGNKMVPKIPASTSS-----SPAISSAQTTNPVTPVVIKQIAVKHVTKNSAT 1683
QY 2118 QSSASOPPRPQOQGVKLTMAQLTQLTQGHGNGQGLTVVIGQGQGTGQ-----LOLI 2169
Db 1684 QSTAS-----SSRVALPLAQI-----KNKLLAQOQOQSTSSSPATSSSPVQKI 1727
QY 2170 PQGVTVLPGGQQLMAAMPNGTVQRFLETPPLATTATTAATTTTSTVSTAA--GTGBO 2226
Db 1728 VSKVNTSTSGOTLQOQVFOGSS-----KLWVGQNAQGVKVIISTSAAOQOQGTSPVQ 1779
QY 2227 QSKLSQPMOVHODKTLPPAQSSS-----VGPAK--AOPQTAQPSARPOQPOQSPAQ 2278
Db 1780 QOGLVOSQPIQSS-----PQOISMTQOQIVVGGORIILSPGQITVITQNVVPSQALQWQO 1835
QY 2279 EVOTQP-----EVOTQTTVSSHVPSEAQTHAQSSKPPQVAASQPSQNSVQOGSPVRV 2330
Db 1836 QIQTOQOQOQHVVVQPOQGFVQSGNIGVQSSPSAQTCLKVLQVVOQSOQSTIEETQIIT 1895
QY 2331 QSPSQT-----RTRPSTPQLSPGQSOQVQVTTTSQPIPIOPHTSLQ-----IPSGOQ-- 2377
Db 1896 TDSNETGTQOVLVNPSTLTAQQLAQGLQVATVNGOQVIVKPLGNNAQQAIVAHIKHGGGN 1955
QY 2378 -----PQSQPOVQSSQTLSGGQTLNOVSVSSPSRPLQIQOQPOP----- 2417
Db 1956 AHIVTSNATVQANPQTSFVKQOALPQSPQOVVQ-----QQQIHQSGPTNFBEGVT 2010
```

```
QY 2418 -----QVIAPVLOQOQOV--LSQIOSQ-----VVAQIOAQOQSG----- 2449
Db 2011 PITQOPVLTAQVAPQAQOQALSVEESLLQNPPTGVIKCVTAQVLTQTEHGPRIVLOGLVG 2070
QY 2450 -----VPQOIKLOLP-----LQIOOSSAVQV----- 2470
Db 2071 NDTAQOOLQVQVQOQLMKQAQESNGKLVGIPKPKIYIYLAQOPENAVQSQPPLTPVHQ 2130
QY 2471 --HQIINVTVQA--ASVQEQLQVRQQLR-----DQOQKKQOQIIEINVNTSKL----- 2516
Db 2131 AAHQOTNNIEIDATLTATTEANSTIKDIAINNGDDOENSKCAETENSITTNESFAGTS 2190
QY 2517 -----LIKVEIIQOVVMKHA--VIEHLKOKKSKMTPAEAREN-----QRMI 2556
Db 2191 SILLEGSHDEPTNLAGLIDISETDLENKQNESFVVTTRYIKSISNALKOGNISPELEKL 2250
QY 2557 VCNQVWKYIILDKTDKEE-----KQAAKKRKRRESVEQK----- 2589
Db 2251 VCMQKQOENANSTNETCSRGVNEALTPSQTDDETKIRTSLRRPNAMTTSSQFNR 2310
QY 2590 ----RSQN-----ATKLSALLFKHKEOLRAEILKRLALLDKDQIQIEVQOELKDLKIK 2639
Db 2311 ILKKNRSKNDEVAELGEOQSLERHKEILKKNILKRSLLERNLQSEIHE-----DVKT 2366
QY 2640 KEDLMOLAQATAVAPCPVTPVLPAAPPSPPPPPGVOHTGLLSTPLPV----- 2693
Db 2367 VQRHVRPLSNA-----SPDEQSENERSG-----EPNLDFKRTEVQ 2401
QY 2694 ----ASOKRKRREEK-----DSSSKSKK 2713
Db 2402 NPHGAGRPKLLFRKKEKLYCICRTPYDDTKFYVGCDCLSNWFHGDVCSITEASK--KLS 2460
QY 2714 KMISTTSKTKTKDKLYCICKTPYDESKFYIGDRCQNTWTHGRCVGIQLOSEALIDYVC 2773
Db 2461 EFICIDCKRARETQOQLYCSCROPYDESOFYICDCKODWPHGRCVGIQLOSEAFIDEYVC 2520
QY 2774 PQOSTEDAMTV--LTPLTEKDYGLKRVLSLQAHKMAWPFLEPVDNDPADYGVYKEP 2832
Db 2521 PECORNDANANMKKLTISNDVEELKNLIKQMLKHSANWFPEPVDKEADYKYVKEP 2580
QY 2833 MDLATMEERVQRRYYEKLTEFVADMTKIFDNCRYNPSDSFFYOCAEVLSEFFVQKLKGF 2892
Db 2581 MDLARMKILESNTYTKLSEFIGNMTKIFDNCRYNPKESFYKCALESEFYVQKIKNF 2640
QY 2893 K 2893
Db 2641 R 2641

RESULT 5
Q9W0T1 PRELIMINARY; PRT: 2669 AA.
AC Q9W0T1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG32478 protein.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Calniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
```



```
Db 631 AASFLNINVVNKKIWAANVARRPSEFAVLLLLFOASLSKVVFANVWHEQLGHTTLQR 690
QY 839 MTSIREBEKEVKVKKKKQ---EBEETMQOATWYKYPFPKHQVWKOKGEBEYRVYTGCGW 895
Db 691 ITSAREERKKLEKREKREDEEERNAFLAENYIKYTLGLKHQVWKOKGEBEYRVHGWG 750
QY 896 SWLSKT---HVRVFPKLPNGTINWYKRSLEGTKNMNDENDESCKRCSRPPKKIKI 950
Db 751 LMLSSRRCGVARRAQPLTHRVVYHTM---GEENDVNEII-----LVDPRTRQF 799
QY 951 EPDSKDEKVGSDAAKAGDQ--NEMDISKIITEKKDQDVKELLDSDSKPCKEEMVEDDD 1008
Db 800 MQOCESSNVGQVCHYLPDQYKNVKVEDVTE-----831
QY 1009 MOTESHVANCQESSQVDVNVNVEGHL--RTSYKKKTKSKLDGLLEIRIKOFTLEEKORL 1066
Db 832 -KIGH-----IDVSKALNAPGRTYISKVARKSRJDDDLDRKLKLAEEV--QMA 878
QY 1067 ERIKLEGGIGIGTSTNSKNLSESPVITRAKEGCGQSDSMROEQSPNANNDQPEDLIQ 1126
Db 879 SKI-----881
QY 1127 CSQSDSSVLMSDPSSHNTKNLYPKDRVLDDVSIIRSPETKCPKQNSIENDIEEKVYSLASR 1186
Db 882 --PSDMKPLLVSSQNTANS-----KOTFLE-----KR 907
QY 1187 GOEPTKSKTKGNDFFIDDSKLASADDIGTLICKNNKKPLIOESDITVSSSKSALHSSVPK 1246
Db 908 LLRLTEVQAKG-----918
QY 1247 STNDRDATPLSRAMDFEGLKGCDSNSTLENSDTVSIQSDSEEDMIVQNSNISBQF 1306
Db 919 -----GPANVNLVNSLAKOIQTVRLQFS-----943
QY 1307 RTREQDVEVLEPLKCELVSGETGNCEDLPLVKGTTEANGKPSQOKKLEERPKNKCSDOI 1366
Db 944 ---QLNRFKAFVRC---YTKEC-----NTNSNAVSQITO-----NFCYSP 978
QY 1367 KLKNTTDDKN---NENRESEKKGQRTSTFQINGKDNPKIYLKGECLKEISESVVSGN 1422
Db 979 CLOKARAKELLLLRKAHTAGSKEIVAAIILGAVKKPSIL-----EOKLTEGK 1028
QY 1423 VEPKVNINKIIPENDIKSLTVKESAIRPFINGVIMEDFNRNSETKSHLLSSSDAG 1482
Db 1029 RE-----STQAVADDSEEGKPAESEAPLDLLQDWEH 1059
QY 1483 ---NYRDSL-----ETLPSK-----ESDSTOTTPPSASCPSNVSNOVEDM- 1521
Db 1060 ARAHAVFSDSLLTCEILVDQECVNTKIKQEVNASSGCNTTPDSNTQDSKDIDYESMD 1119
QY 1522 ---BIETSEVKKVTSSPITSEESNL-----SNDFID---ENGLP 1555
Db 1120 VCSNVEIESTEDSIVTGLSNGNAEDVMDPPGWRKRNRKQSKSKSYIGTKDVLDTLDKIDP 1179
QY 1556 INKNENVGESKRKTIVITEVYMTSTVATVETSKIVKVEGDKQTVVSSSTENCAKSTVIT 1615
Db 1180 LNK-----QNRFPFIT---ARPVKREC-----VKKYRETFENGNERV- 1214
QY 1616 TTVTVTKLSTPSTGGSV---DIISVKEQSKTVVTTVTVDLSLATTGTTVTSMTVSKYST 1671
Db 1215 -----YSTSSPGRVYLLNDAAKLYEA-----YKT 1240
QY 1672 RDVKLKMFSRPPKKTSGTALPSYKRVFTYKTSKISFVLPNDOLKKLARKGGIREVPYN 1731
Db 1241 EDKSTITK--KPSYSR-----YPLISNFLTTHKKRSLVLPRLPELLKLARLGKSSNTGNFH 1294
QY 1732 YNAKPALDIWYPSPPTFCITWRVLOTVKSLAGVSLMLRLIWLASLRDMDMAAKVPPGG 1791
Db 1295 HAAKNN-TWQYQCSRPLFTWCYSYRTSNATSLUSSLALQRLIWLSCLRDMDMIAK--PPST 1352
QY 1792 GSTRTTSETFEITTEILIRRDVGPGYRIFCYCIRKIIICPGYPTPKETPTPQRGLRS 1851
Db 1353 DGKHQVTTDTEIYVTELLKLRHSGRYGEKTSYLRKRKVVIPLEMPKTVREV-TSIRSGLR- 1410
```

```
QY 1852 SALRPKRPTPKOTGPVIIETWVAEBELELWEIRAFARVERKEKAQAVEQ--AKKLEQO 1910
Db 1411 ---KRRAESPQTEPOIITEWVDEKLELWEIKFMGEKARLSAATRSVASRQLE-- 1465
QY 1911 KPVIATSTSPSSSTSTISPAQKVMVAP-IGSVTTTGTWKVLLTTKVGSPATVTTQONK 1969
Db 1466 ---ASGNSGNTSTNGALVAGRVOLAPKLSEDKV--EKMBEQQLKL--ORAVHQOR 1515
QY 1970 NFHOTFATWVKQGSNSGVVOQKVLG---IIPSTGTS---QOTFTSFQ---PRTA 2018
Db 1516 ---LVATGEITRSVTPVKGVIGSRVIVKNPDGTTRIIOQAIVTQVSRGTGANTA 1567
QY 2019 TVTIRPNTSGSGTTSNS-----QVITGPQ-----IRGPMVIRTP---LQ 2057
Db 1568 AAAASPVTGSGTSTOSNPSTSTPHKVQIIRGPDGKVSRLGNPCQOLVQMPDGKLHVLTT 1627
QY 2058 STIGKALIRPVAVQCAPQOVMTIIRGQPVSTAVSAPNTVSTPGCKSLTSATSTSI 2117
Db 1628 TTSNSAGOGNKKVPKIPASTSS---SPAISAQTTTPNPVTPVVIKQIAVKHVTNSAT 1683
QY 2118 QSSASQPPRPQOGQVKLTMAQLTQTOGHGNGOGLTVYIOGOGTTGQ-----LQ 2169
Db 1684 QSIAS-----SSRVALPLAQI-----KNKLLAQOQOQSTSSSPATSSSPVKI 1727
QY 2170 PQGVTVLPBGQOQLMQAMPNGTVORFLPTLATTTATTTTSTTAA---GTGEOR 2226
Db 1728 VSKVYNTSTSGOTLQQVFOVSGS---KLVGQNAQOGKVIISTSAQOQGTSPVQ 1779
QY 2227 QSKLSPQMVQHOHDKTLPPAOS--SVGPAKAO-----PQ 2258
Db 1780 QOQLVQSOPIOS---POQISMVQVGNQTKVQIQIVNTSNVQOQIVVGGORILLSPG 1835
QY 2259 TAQPSARPQOTQPSQAPEVOTQP-----EVOTQTVSSHVSESAQTHAQSSKP 2310
Db 1836 QTVTQVNPQSOALQWVQOQIQOQOQHHVQPOQOQFVQVQSNQIVQSPSAQTKLVK 1895
QY 2311 QVAASQPOSQVQSPRVQSPST---RIPSTPSQLSPQSQSOVOTTTTQPIQ 2365
Db 1896 QLVVQOQSOQTIEKTOITTTTDSNETGTQOVLVNPSTLAOLAQGLQVATVNGQQVIVK 1955
QY 2366 PHTSLQ-----IPSQO-----POSQPVQSOVSTLSSGOTLQVSVSSP 2405
Db 1956 PLGNNAQIVAHIKHQDGNNAHIYVNSATVPAQNPQTSVPKQOALPPQSPQVQVVO-- 2013
QY 2406 SRPOLQIOQPOP-----QVIAVPOLOQOQVQV--LSQIOSQ-----V 2439
Db 2014 ---QOQIHQOQSPTNFESGVTPITQOQVLTQAVQAPAOQOALSVEESLLQNPQPGTVIKCV 2070
QY 2440 VAOIOAQOQSG-----VPOQIKLQLP-----IQ 2461
Db 2071 TAQVLQTEHGPVILQGLVGNDFTAQOQLQVOTQVKOOLMKAQESNGKLVGLGPKIYLA 2130
QY 2462 IQOQSAVOT-----HQIQNVVTVQA---ASVBOLOQRVQOQLR-----DQOQK 2500
Db 2131 VOPENAVQSQPPLTPPVHOSAAHQOTNNIEADATLTATTYEANTSTIKDIANNNGDOENS 2190
QY 2501 KQOQOIEINVTPSKL-----LIKVEIIQOVVMKHNA--VIEHLKQK 2541
Db 2191 KCAETENSNIITTNESFACTSLLGESEHDEPTNLAGLDISETDLENKQNESFVTVIRGVIQ 2250
QY 2542 KSMTPAEREN-----QRMVNCQVMKYILDKIDKEE-----KQAAKKRKEESVEQ 2588
Db 2251 KSISNALQOGLNSPLEKIVCMQKQOENANSTNEWETCSRGSVNEEALTPSRQDDEW 2310
QY 2589 K-----RSKON-----ATKLSALLFKHKQLRAELIKKRAL 2619
Db 2311 KIRTSILRRPNAMTTSSQFNRIILKKNRSKNDEVAELGEQKQSLERHKLLKNILKRSL 2370
QY 2620 LDKDLOIEVQEBELKDLKIKKEDLMQLAQATAVAAPCPVTPVLPAPPAPPPPPPG 2679
Db 2371 LERNLOSEIHE-----DVTKTVQRHVRPLSNA-----SPDQSE 2404
```


Db	882	--PSDMKPLLYSSQNNTANS	-----KQTFLE-----	KR	907
Qy	1187	GOEPTKSTKNDFFIDDSKLASADDIGTLICKNKKPLIQEESDTIVSSSKSALHSSVPK	1246		
Db	908	LLRLTEVOQAG	-----	918	
Qy	1247	STNDRATPLSRAMDFEGLGCDSESNSTLSSSDTVSIQDSSEDMIVQNSNESISEQF	1306		
Db	919	-----GPANVULEVNSLAKQITQVRLQFS	-----	943	
Qy	1307	RTREQDVELEPLKCELYSGESTGNCEDRLPVKTEANGKPKSOQKLEERPVMKCSDOI	1366		
Db	944	-----QLNFAKVFRC-----YTRKEC-----NTNSNAVSOITQ-----	NTCYSPL	978	
Qy	1367	KLKNTTDKN	-----NENRESEKKQORTSTFOINGKDNKPKIYLKGECLKEISESRVSGN	1422	
Db	979	CLQKARAKKELLRLKRAHTAGNSKETVAAILGAVKKPSIL	-----EOKLTEGK	1028	
Qy	1423	VEPKVNNKIIPENDIKSLTVKESAIRPFINGDVIEMEDFERNSESTKSHLLSSDAEG	1482		
Db	1029	RE	-----STQVAVDDSEGRPAESEAPLDLLLODWEH	1059	
Qy	1483	-----NYRDSL-----ETLPSK-----ESDSTQTTTPSASCPSNSVNOVEDM	1521		
Db	1060	ARAHAVPFSDSLTECILLVQBCVNTKIKQEVNASSGCNTTPDSNTQDSKIDYIESMD	1119		
Qy	1522	-----ELETSEVKKVTSPITSEESNL	-----SNDFTD-----ENGLP	1555	
Db	1120	VCSNVEIESTDSIVTGLNSGNAEDVMTPGWRRRNOKSKSYIGTKDVLQDLODKDIP	1179		
Qy	1556	INKNENVGESKRKVTIETVTMTSTVATESKTIVKEGDKQOTVYSSSTENCAKSTVTTT	1615		
Db	1180	LNK	-----QNRFPIT-----ARPVKREC-----VKKYERETPENGNERV	1214	
Qy	1616	TTTVTKLSTPSCGGV	-----DIIISVKEQSKIVTTVTVDLSLTGTLVTSMTVSKEYST	1671	
Db	1215	-----YSTSSPRGRVYLLNDAAKLYEA	-----VKT	1240	
Qy	1672	RDKVKLMKSPRKTRSGTALPSYRKFTKSTKKSIFVLPNDLKKLARKGIREVPYFN	1731		
Db	1241	EKSTITK--KESYR	-----YPLISNFLTHKKRSILLVLPRELLKLARLGKSKSTNGFH	1294	
Qy	1732	YNAKPAIDIWPPSPRPTGIIWRYRLQTVKSLAGVSLMLRLLWASLRLWDDMAAKVPGG	1791		
Db	1295	HAAKNN-TIWQOCSRPLRPTCWSYRTSNATSLSSALQRLILWCLRWDDMIAR-PPST	1352		
Qy	1792	GSTRTETSETITELIKRRDVGPIRFEYCIRKICPIGVPTPTKETPPORKGLRS	1851		
Db	1353	DGKHQVTTDTEIVTLELLKLRSGRYGKTSYLRKRVIPLEMPKTVREV-TSIRSGLR	1410		
Qy	1852	SALRPRPETPKOTGPVIIETWVAEELELWIRAFARVEKEKAQAVEQO-AKKRLBQO	1910		
Db	1411	---KKRAESQPTEQITEEWDKLELWEIFKFMGEQEKARLSAVTRSVASRQLE--	1465		
Qy	1911	KPTVIATSTSPSTSTISPAQKVVAP-LSGSVTGTGKVLTKVGSPTATVFOQNK	1969		
Db	1466	---ASGNGSNTSGALGVQAPKLSQEDYK--EKMEQOQLK---QRAVHQORR	1515		
Qy	1970	NFHQTATWVQKQSGVGVQOQVLG	-----IIPSTGTS---QOTFTSFQ-----PRTA	2018	
Db	1516	-----LVATGEITRSTVPKQGVIGSRVIVKPNPDGTTIRLIQAQVQTSRGANTA	1567		
Qy	2019	TVTIRNTPSGSGTTSNS	-----QVITGPO-----IRPGMTVIRTP	2057	
Db	1568	AAAAPTGVGSGTSTQSNPSTSTPHKVOIIRGPDGKVSVRGLNPQOOLVQMPDGKHLVJTT	1627		
Qy	2058	STLGKALIRTPVWQVQAPQOQVMTQIIRQPVSTAVSNAPTYSSTPGQKLSASATSN	2117		
Db	1628	TTSSNSAGQGNKKMKVPIKPASTSS	-----SPAISSAQTTTNPVTPIQJAVKHVTKNSAT	1683	
Qy	2118	QSSASQPPRQOQKILTMALQTLQTHGGNGGLTWVVIQGGQQTGTQ	-----LQLI	2169	
Db	1684	QSTAS	-----SSRVALPLAQI-----KNKLLAQOQOQSTSSSPATSSSPVQKI	1727	

RESULT 7

Q8VDN7

ID Q8VDN7

PRELIMINARY;

PRT;

645 AA.

Db 1002 ----- 1001
Qy 1502 TTPSASCPESNSVNOVEDMEIETSEVKVYKVTSSPITSEESNSLNDIFENGLPKNNEN 1561
Db 1002 -----SG 1003
Qy 1562 VNGESKRKVTITVTMTSTVATESKTVIKVEKGDQTVVSSSTENCAKSTVTTTTVTYK 1621
Db 1004 VLGEK----- 1009
Qy 1622 LSTPSTGSGVDIISVKEKSTVVTVTDTSLTTGGTLVTSMTVSKEYSTRDKVKLMKFS 1681
Db 1010 ----- 1009
Qy 1682 RPKKTSRGTALPSRYKVTSTKTSIFVLNDDLLKLARKGGIREV--PYFNYNAKPAID 1739
Db 1010 -----AWPIPEIQTSTKRGKSIFFVLQKTLRQIMGGCQOVVMPGFSAGIKSNLL 1062
Qy 1740 IMPYPSRPFTFGITWRYLQTVKSLAGVSLMLLWASLWDDMAAKVPPGGGSTRTETS 1799
Db 1063 IMPYAPRPFLDLCWKQTLNARSLHAVALQLKIWSIKFNEFD---PDTHPRRVVI 1119
Qy 1800 ET--EITTEIIRRDVGPIRGFEYCIIRKIICPI-GVETPKETPTPORKG-----LR 1850
Db 1120 DTPSHDRRLIRHKEMPPYGOYERYEMEIEIIPLYDEPEEDESLSRNRGSSSEFSHR 1179
Qy 1851 SSALRPKRPT-----PKOT-----GPVITWVAEELELWE----- 1883
Db 1180 SSARKRKQHEFLSLKFGNPKKSNAPRSLNRRATAIRREWVDGVTLKVEIKDYW 1239
Qy 1884 --TRAFAERVEKEKAQV-----EQAKKRLQOKPTVIATSTSTPSSTSTISPAQ 1934
Db 1240 KWTRAEAKTKRKLKLEATRAKAKADEERRRQQOQORSVARIPV-PMHS-----LIPSE 1294
Qy 1935 KVMVAPITGSGV---TTGKMWLTTKVGPATVTEQON-----KNEHGTATATWKQOG 1983
Db 1295 RNNV-PYLSGQOORRNGNERGLEKYNSSSVSPAAGVASTPPGYPHQPQNIIRQAG 1353
Qy 1984 SNGGVQVQOKVLGIIPSSGTSTGQFTSPQPRATVTRPNTSGSGGTTSSNOVITGPQ 2043
Db 1354 YNQ-----LPRKPTTSPFNFSQ-SP-VATITPTTQLRAAGADG---VVRVAVM 1396
Qy 2044 IRPG-----MTVIRTPLOQSTGLKALIRTPVWOPGAPQVMTQIIRGO----- 2087
Db 1397 MTPGNKSTVNTNSTPYQ-----ALNRQYQLQROQOQPAVRRLTNGYHFMGTMRGGGR 1451
Qy 2088 -----PVSTAVSAPNTVSTPGKSLTSAT-----STSNLQSS 2120
Db 1452 NPSVQMHQRLPQNRALORPFGESTTEMRRVTEAIPDNDGDQPPVPIPRDFTSNFDAQ 1511
Qy 2121 ASQPPRQOGOVKLTMAQLTQLTQGHGNGQLTVVIGQGQTTGLQL-----IPQGVTVL 2176
Db 1512 RAQOOHPQSRPVYSTPAQIMRTTPGGVKH--NVILMKASDGTQKMWLPKPGQFPPTVI- 1568
Qy 2177 PGQOQLMQAMPNGIVQRFELFTPLATTATSTTTTTVSTTAAGTGEQRQSL----- 2230
Db 1569 -STGQRVVPYRQPTAVQOQOLYT-----ATPGTRVVRIPNANGAPRQODHGVMRVVQ 1621
Qy 2231 --SPQMVQHDK--TLPPAQ-----SSSVGPAKAQPTQAQSPARPOQTQPOS 2274
Db 1622 ASQPRAMEYDDGTPPGQOQRYVLOGNSGTPNPNPKVSSRGPRGLTWMQVQOQQ 1681
Qy 2275 PAQEVQTOPEVOT---QTTVSSHVPSEAQ 2301
Db 1682 QHNPAYDMPDATTGFAVSTTTTQVPEEQ 1711

RESULT 10

ID Q95Z08
AC Q95Z08; PRELIMINARY; PRT; 1713 AA.
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE F26H11.2b protein.
GN F26H11.2 OR F26H11.2B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
DR EMBL; Z81515; CAC42289.1; -.
DR WormPep; F26H11.2b; CE28009.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00384; AT_hook; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
SQ SEQUENCE 1713 AA; 196824 MW; 178E470017D9AD5A CRC64;

Query Match 5.7%; Score 848.5; DB 5; Length 1713;
Best Local Similarity 17.4%; Pred. No. 5.8e-28;
Matches 434; Conservative 295; Mismatches 619; Indels 1147; Gaps 74;
Qy 4 EEEEE---DGAETQDSEDD---EDMEEDDDSDYPMEEDD---DDASCYTE 52
Db 169 KROEEDIYDESEDESESESSDDFMLENDQVVOEEELNLTDIKEKGLDEENKVC-- 226
Qy 53 SSFRSHSTYSSTPGRKRPRVHRPRSPILEE--KDIPPLEPKKSEDLMPVNEHIMNVIAI 110
Db 227 -----PWLDEDPASLPKLELPSESSQDIPITASTMDAVEI 261
Qy 111 YEVLRFNFGTVLRUSPRFEDFCALVQSQECTMAEMHVLLKAVLREEDTSNTFPAD 170
Db 262 YEILRSYHRLRITPTTFEDFCALISHNNSCIMAEVHALLNRLCKLSDDEEQTHYSVTE 321
Qy 171 LKDSVNSTYFIDGMWPEVLRVYCSDKHHVLPYQEA-----EDYPGPVENKI 222
Db 322 TNSVNMIMHMDTLTYAEILROYIEA-----YPFADASVRDAINVDNYPFGVDANI 374
Qy 223 K-----VLQFLVDQFLTNTIAREELMSEGVIOYDHCRCVCHK-LGDLIC 265
Db 375 QRDSEFFFNKHKVFLRVLLFMSYRFLYSSEFKKLVNNVCKFQDENCRCVCKSSGRVVG 434
Qy 266 CETCSAVYHLECVKPPLEEVPEDEWCEVCAHK-VPGVTDCAVETQKNKPYIRHEPIGY 324
Db 435 CTQCEAAHFVCE--SHLKPPPE-VLCNICKKNSAVRGVLPPEAVDREP--LRSQPIGR 489
Qy 325 DRSRRRYWFLNRLIIEEDTENENKIKWYSYKVLAEILCDLKDXYWAEKLCILEEM 384
Db 490 DRYGRYWFIVRLVYQSLDDETE-----LYYSTVPQYQLQKLDYTYEKDLCDTIRLR 545
Qy 385 REEIRHMDITELTNKARGSNKSFALAAANEELISIRAKK---GIDNVKSPETEKKOK 441
Db 546 IDEFLQMALTVEMTSERR-----EAALETWVKQLIGYDFAEATTQ----- 588
Qy 442 NETENDSKDAKNREEFEDQSLKSDDKTPD---DDPEGKSEETFEVGDGKNSVANL 498
Db 589 -----IYLHRDSMKRMASILRDCAKGQVKQEVKL----- 618
Qy 499 GDNNTNATSEETSPSGRSPVGLSETPDSSNNAEKKVASSELQDYPPEEPNKTCSNNTS 558
Db 619 -----EE--PVGQSPVKCVQFVEDSI-----LPE----- 641

```
Qy 559 ATTTSIQPNLENSSSSELNSQSESASAKAADDPENGRESHTPVSIQIEIVGDPFTSEKST 618
Db 642 -----SMIGIFADKLINTFWSSGA-----TOEELVEQV-----DI 672
Qy 619 GELSESPGAGKASGSTRIITRLRNPDSKLSQASQOAAAAHANKLFREGREVLVNS 678
Db 673 SDNFDAPSA-----681
Qy 679 QGEISRLSTKKEVIMKGNINNYFKLQEGK---YRVYHNOYSTNSPALNKHQHRHEDHDKR 735
Db 682 -----NLWRMGDEGNDQTFMTYNYYSRNMSESEFLTRKKAADKK 721
Qy 736 RHLAKHFCULTPAGEFKW---NGSVHGSKVLITLTLTITOLNENIPSSGFHFHNWASHR 791
Db 722 KYMASKF---AQIDNFDVVAKNROFYGDASHLCHFIMWTLOOVILKNIPIDLHRRKWPEFA 779
Qy 792 ANWIKAVQMSKPREFALALAILCAVAPVVMPLPIWREFLGHTRHLRMTSIEREKEKVK 851
Db 780 KGFDELEVSVADDYKKLVTCLLKLDCAVRKTIFFPQWNGLGQTRLERITVDORENFMKEQ 839
Qy 852 KKEK-----QBEETMQOATWVKYTFPVKHQVWQKGEYRVTG---YGGWSWLSKT 901
Db 840 QRLKKIDADALTOKLDDSFVRVNYMKPKWNTY-ILRQGETYRNAGKSGMGWAWAAK 898
Qy 902 HVYREVPLPGWNTVNYRKSLGKTNMNDENDESDDRKCSRPKIKIEPDSKDEYKG 961
Db 899 VYEKI-----QVPESKPLPLAVTEEIKTESVNRK-----930
Qy 962 SDAAGADONEMDISKITEKKQDVKELDDSDSKPKEEPMWDDDMKTESHVNCRESS 1021
Db 931 -----ARRLELLVSKIT-----942
Qy 1022 QVDVVNVSEGHILRTSYKKTKKSSKLDGLLERRIKQFTLEEKQRLKIEKGIGKIGKT 1081
Db 943 -----KKQRSG-----GKS 952
Qy 1082 STNSSKNLSSEPVITKAREGQSDSMRQEQSPNANNQPEDLIOGCSQSDSVLRMSDPS 1141
Db 953 SKKPTFELT-----NGCYSPPSCRS-----971
Qy 1142 HTTNKLYPKDRVLDDVSTRSPETKCPKNSIENDIEEKVSDLASRGQPTKSKTGNDFF 1201
Db 972 -----NPNRK-----YSP-----980
Qy 1202 IDDSKLAGADDIGTLICKNKKPLIOESDITVSSSKSALHSSVPKSTNDRDPTLSRAMD 1261
Db 981 -----MCRN-----984
Qy 1262 FEGKLGCSSESTLNSSDTVSIQDSSEDMIVQNSNESISEQFRTREQDVEVLEPLKC 1321
Db 985 -----984
Qy 1322 ELVSGESTGNCEDRLPVKGTANGKKPSQOKLEERPYNKCSQDIKLKNTDDKKNENRE 1381
Db 985 -----GYLSAKOAHDERKLEE-----1001
Qy 1382 SEKKGORTSTFOINGKDNPKIYLKGECLKEISERVVSGNVGPKVNNINKIIPENDIKS 1441
Db 1002 -----1001
Qy 1442 LTVKESAIRPFINGDVIMEDFERNSETSKSHLSSDAEGNYSRDSLETLPSTKESDSTQ 1501
Db 1002 -----1001
Qy 1502 TTTPSASCPESNVQVEDMEIETSEVKVTSSPITSBEESNLSDNDFIDENGLPINKNEN 1561
Db 1002 -----SG 1003
Qy 1562 VNGESKRKTVITEVTTMTSTVATESKTVIKVEKDGKQTVSSSTENCAKSTVTTTTTTVK 1621
Db 1004 VLGEK-----1009

1622 LSTPSTGGSDIISVKEQSKTVVTTVTDLSLTGTLVTSMTVSKEYSTRDKVKLMPKS 1681
1010 -----1009
1682 RPKTRSGTALPSYRKFVTKSTKKSIFVLPNDDLKLLARKGGIREV--PYFNTNAKPAID 1739
1010 -----AWPIPETOTFSTRKGGKSIFVLOKKILROMIMGGCCQVYMPGFSAGIKSNLL 1062
1740 IWPPSPRPRTGTIRYRLQTVKSLAGVSLMLRLLMASLWDDMAAKVPPGGGSTRTETS 1799
1063 IWYPAPRPDLCLCWKQTLNARSLHAVALQKLIWSSIKFNEFD---PDDTHPDRRVVI 1119
1800 ET--EITTEIIRRDVGPYGIIRPEYCIIRKIIPI--GVPTPKETTPQKRG-----LR 1850
1120 DTPSHDRRIIRHKEMPPYQYERYEMEIEIIPLYDEPEEDESLSNRGGSSEFSHR 1179
1851 SSALRPKRPE-----PKOT-----GPVIEIETWAEELWE-----1883
1180 SSSARKRQRHHEFLSLKFGNYPKKSNFRSLDNRATAIRREWVDGVTLYKVEIKDYW 1239
1884 --IRAFAEVEKEKAQAV-----EQAKKRLEQQKPTVIATSTTSPTSTSTTSPAQ 1934
1240 KWIRAEAEKTRKLEATRAQAKADEERRRIQQOQORSVARIPV--PMHS-----LIPSE 1294
1935 KVMVAPISGSV---TTGTRKMLVTKVGSPTVTFQON-----KNFHQTEATWVKQG 1983
1295 RNVV--PYLGSQQRRPNNGERGFLEKYNSSVSPQAHGYASTPPPGYHQPNIIKQAG 1353
1984 SNSGVVQVQKVLGIIPSSGTSTQOTFTSFQPTATVIRPNTSGSGGTTSSNOVIITPQ 2043
1354 YNQ-----LPRKPTTSPFNQOS--RP-VATITPTQLRAAGADG--VVRVAM 1396
2044 IRPG---MTVIRTPLOQSTLGLKAIIRTPVWVQGPAPQVMTQIIRG-----2087
1397 MTPCNKSTVNTNSTPYQO-----ALNRQYQLOQOQOQPAVRRLTNGYHFMDGTMRGGGR 1451
2088 -----PVSTAVSAPNTVSTPGOKSLTSAT-----STSNIQSS 2120
1452 NPSVQMHQRLPONRAALQORPFGESTEMRMVRTEAAIPDNDEQOPPVIPRYDPTISNDAQ 1511
2121 ASQPPRPQOQGVKLTMALQTLQTOGHGNGOGLTVVIOGQGTQLQLQ-----IPQGVTVL 2176
1512 RAQOQHHPQSRPVYSTPAQIMRTTQPGVKH--NVILKASDGTQKMWLKPQGPPTGTVI- 1568
2177 PGQGOQLMAAMPNGTVQRFLETPLATATASTTTTTTSTAAAGTGEORQSKL-----2230
1569 -STGQVVVYRQPTAVQOQRLYT-----ATPGTRVVRIPNANGGAPRQODHVMRRVQ 1621
2231 --SPQ-MQVHODKTLPP-----AQSSSVGPAKAPQTAQPSARPO-----PQTQ 2271
1622 ASGPRAMEYMDQGTTPPGQVRYVLOGNSGTENVAPPKVSSRSGRGGGLTQMOWQOQ 1681
2272 PQSPAQ-----PEVQTOPEVOTOTVSSHVSEAQ 2301
1682 QHNFQAHYMDPDDATGFAVSTTT---EQVDPDEQ 1713

RESULT 11
Q960Y3 PRELIMINARY; PRT; 1022 AA.
ID Q960Y3
AC Q960Y3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE LD30146p.
GN E(BX) OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
```


RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 FN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Bantz J., An H., Baldwin J., Bantz J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Matel B., McShrefi A.,
 RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 FN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 FN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 FN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF003808; AAF58087.2; -
 DR FlyBase; FBgn013988; Strn-Mlck.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR007110; Ig-Like.
 DR InterPro; IPR003598; Ig-C2.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 13.
 DR SMART; SM00409; Ig; 21.
 DR SMART; SM00408; Igc2; 20.
 DR PROSITE; PS00835; IG_LIKE; 20.
 DR PROSITE; PS00290; IG_MHC; 1.
 KW Immunoglobulin domain.
 SQ SEQUENCE 7210 AA; 808965 MW; 6B2E7395C07140D2 CRC64;
 Query Match 4.3%; Score 641; DB 5; Length 7210;
 Best Local Similarity 18.4%; Pred. No. 2.3e-18;
 Matches 601; Conservative 538; Mismatches 1187; Indels 938; Gaps 135;
 QY 5 EEEEDGDA-----EETQDSD-----DEEDMEEDDDSDYPPEMEDDDDDASVCT 51
 DB ESQNKDAAGDIKKISETDYYDHSIEKKIEPRKSRKDLDFEELEE-----L 2618
 QY 52 ESSFRSHSYSTPGRRKRVHRPRSPILEEK-DIPPLEFFPKSSDLMPNHEIMNVIAT 110
 DB 2619 KASAKKQGDQIEQKSQKEV-----SEVVAEKISEGKEEPPKPEM----- 2661
 QY 111 YVLRNFGTVLRSPRFDFCAALYSQBCQTLMAEMHYVLLKAVLRREE--DTSNTTGP 168
 DB 2662 -----DTEAKSEKATVLDKQVLEKELEASAEKQGD 2692
 QY 169 ADLKSVDNSTLFIIDGTMFWPLRVYCESDSKEYHHVLPQAEADYPYGVENKIRVQL 228

Db 2693 QDVEKSKS-----PEVSEVVAEKISE-----ETIEPKKP----- 2723
 QY 229 VQDFLTNTIARELMSEGVQIYDDHCRVCHKGLDCCETCSAVYHLECVKPPLEVPED 288
 DB 2724 -----EVKOTEIKSEKATALD-----KQVLEKELE 2749
 QY 289 ---EMQCEVCVAHK---VPGVTDCVAB-----IQKNKPYIHREPIGVDRSRKRWFLNR 336
 DB 2750 ASAQKQCDQDVEKSKQKPEVSEIVAIEKISEKTEIEPKKPEVKDTEIKSEKATA-----LDK 2805
 QY 337 RLIIEDTE-----NENEKKIWTYSKVLQALIEDCLDRYWEALCKLILEEMREIHR-H 391
 DB 2806 QVLEKELEASAQKQDQDVEKSKQKPEVSEV-----AE--KISEETIEPKKPE 2854
 QY 392 MDITEDLTNKGNSKFSLAAANEETLESIRAKGDDID----- 429
 DB 2855 VKTEVSKSEKATVLDKQVL---EKELEASAQKQDQDVEKSKQKPEVSEVVAEKISEET 2911
 QY 430 -----NVKSPETEKKDKNTEKNDSDAKNREEFEDESLEKDS----- 467
 DB 2912 IEPPKKPEVKDTEIKSEKATALDKQVLE--EKELEASAQKQDQDVEKSKQKPEVSEV 2969
 QY 468 -----DDKTPDDDPQGGKSEEPTEVGDGNSVSANLGDNTTNTATSEETSPSEGRSP 518
 DB 2970 EKISEKIEPKKPEKTEVSEKAT-VLDKQVLEEKEL-----EASAQKQDQD 3019
 QY 519 VGLSETPSSNMAEKVASELPQDYPEPN-KTCESSTNTSATTTSIQNLNSENSSSEL 577
 DB 3020 VEKRSQKPEVSEVVAEKV-SEGKIEPKKPEVKTEAKSEKATTLDMQVLERELEASAQ 3078
 QY 578 NSSQSSEKAAADDPENGE-----RESHTPYVSIQIEIV----- 609
 DB 3079 KQGDQDVEKSKQKPEVSEVIAEKISEKIEPKKPEKTEVSEKATVLDKQVLEEKEL 3138
 QY 610 -----GDTSEKST--GELSEPGAGKASGSTRITRLRNPDSKLSQKSSQVAAAA 660
 DB 3139 EASAQKQDQDVEKSKQKPEVSEVVAEKVSEK---IEPKKPEVKTEVSEK-----A 3190
 QY 661 HEANKLFKEGKV-LVNVSG-----EISRL--STKKEVIMKGN-- 696
 DB 3191 TTLDKQVLEKELEASAQKQDQDQKSRDDIITLKERLETSEKALGSSVDEILRESREI 3250
 QY 697 INNYFKLGQSGKYVYHNOYSTNSFALNKH-----QHREDHDKRHLAHKFCLTIPAGEF 750
 DB 3251 VNNL-----EDDKVVAKH-----LFKLRDHIVITYDGKRGEEKKELEPESFI-----EL 3295
 QY 751 KWNGSVHSGKVLITLTLTITOLENNIPSSFFHPNASHRANWIKAVOMCKPREFAL 809
 DB 3296 LCEASPEAAEKVKNLYK---EIKTNVILT-----RATIOLIDDSNMF-- 3335
 QY 810 ALATLECAVXPVMPVLPWREFLHTRHRTMSTEREKEKVKKK---EKKQEEETMOQA 866
 DB 3336 -----TKPSLLIP-----KLINLER-VAVKIQSETVYVDRSEKMSISLQQS 3374
 QY 867 TWKXYT-----FPVKHGVKQGEYRVTVGYGWSWISKTHVYRFVKPLPGNTNV---NY 918
 DB 3375 LMDIFVLDDPDDTEVLEPKPIENIKITLLSDYIEKKDGPLLTAVINGKNVYSQHI 3434
 QY 919 KSLGKTNNDNNDNDSKRSRSPKKIKIE--PDSEKDEVKGSDA-----AKGADQNE 972
 DB 3435 LTIIIEVK-QLTENHDQKEDVSNADNFADEKREESQKEEIKDSEAKHKSKVSEKKS 3493
 QY 973 MDISKTEKKQDQVLEKLDSDSD----- 995
 DB 3494 IEEEKLEDDKKEKQTESAIDSEKSAEYSEIVSEKITDEKAQESOKKEVSEAKPKKAV 3553
 QY 996 ---KPCKEEPEVDDMKTESHVNCQESSQDVVYN-VSEGF---HLRTSYKKTKTSKSLD 1048
 DB 3554 LEKSIIEEKLDEKKEKQTESAID-EKSOKAEVSEIVSEKITDEKAQESOKKEVSEKKS--- 3609
 QY 1049 GLLEPRRIKQFTLEEKQRLKLETKLEGIGIKGKTSTNSKNLSSEPVITKAKGCGQSDMR 1108

```
Db 3610 ----EAKPKKAKVLEKKSIEEAKLE-----DKKETQDTSADIDEKSQAEEVSE-TVSEKIT 3659
Qy 1109 QEOSNANNDQPEDLIGCCSQSDSSVLRMSDPSTHTNKLPKDRVLDVDSIRSPTKCPK 1168
Db 3660 DEKAQESQKEEVKD-----SEAKPK-----KAKVLEKKSIEEKLDDKK 3698
Qy 1169 QNSIENDIEBK-----VSDLAS-----RQOETPKSKTKGNDFIDDSKLASADDITGLI 1217
Db 3699 EKOTESAIDEKSQAEEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIEEK 3758
Qy 1218 CNKKPLIOBESPTIV--SSKSALHSSVPKSTNDRDATPLSRAMDFEGKLGCDSPSNST 1275
Db 3759 LEDKK---ETQDTSADIDEKSQAEEVSEIVSEKITD-----EKAQESQKEEVKDSEAKPK 3809
Qy 1276 LENSSTVSIQSDSEEDMIVONSISIQFRTREQDVEVLEPLKELYSGESTGNCEDR 1335
Db 3810 KAKVLEKKSIEEKLDDKKEKOTESAIDE-----KSQAEEVSE-IVSEKITDEKAQESQKK 3864
Qy 1336 LPVKGTGANGKPKSQOKLEERP--NKCSDQIKLNT---TDKKNENRESE-----1383
Db 3865 -EVKGEA---PKKAKVLEKKSIEEKLDDK--KEKOTESAIDEKSQAEEVSEIVSEKIT 3919
Qy 1384 -KKQRTSTFQINGKDNKPKIVLKGCLKEISBSRVVSGNVPEKPVNNIKIIPENDIKSL 1442
Db 3920 DEKAQESQKEEVKDSEAKPK---KAKVLEKKSIE-----EKLKKEKOTESAIDEK 3969
Qy 1443 TVKESAIRPFINGDVIMEDFNERNSETSKSHLSSSDAEGNYRDSLE--TLPSTKESDST 1500
Db 3970 SQAEEV-----SEIVSEKITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIEEKLDDK 4023
Qy 1501 QTTTPSACPENSV-----NOVEDMEIETSEVKKVTSSPTTSEENSLNDFIDENG 1553
Db 4024 EKOTESAIDEKSQAEEVSEIVSENITDEKAQESQKKEVKDSEAKPKKAKVLEKKSIEEK 4083
Qy 1554 LPINKNENNGESKRKTIVETVMTSTVATESKTIVKVEKGDQKQVWVSTENCAKSTV- 1612
Db 4084 LEDKKETQDTSADIDEKSQAEEV-----SEIVSEKITDEKAQESQKEEVKDSEAKPKKAKVL 4139
Qy 1613 ----TTTTTIVTKLSTPGSGVDIISVKEQSKTVVTVTTVDSLTSTTGGTLVTSMTVSEK 1669
Db 4140 EKKSIEEKLDDKKEKOTESAIDEKSQAEEVSEIVSENITDE-----KAQESQKK 4189
Qy 1670 STRD-----KVKLM-----KFSRPKTRSGTALPSYRFVTKSTKSFVLPNDLL 1715
Db 4190 EYKDSAKPKKAKVLEKKSIEEKLDDKKEKOTESAIDE-----KSQAEEVSEIVSEKI 4243
Qy 1716 -KKLRKGIRPEVPEVNYNAKPAIDIWYPSPRPTGIVRWVRLQVTKSLAGVSLMLRL 1774
Db 4244 TEEKAESQKEVK--DSKAKPK-----KAKVLEKKSIEEAKL-----4279
Qy 1775 WASLRWDDMAKVPVPGGSTRTETSETITTEIIRKRDVPGYIRFEYCIIRKIICPIGV 1834
Db 4280 -----EDKKET--QDTSADIDEKSQAEEVSE-----EIVSEKI-----4308
Qy 1835 PETPKETPTPORKLSSALRKPRTPKOTGVIITWVAEELELWEIRAFARVEKE 1894
Db 4309 --TDEKAQESQKEEVKDSEAKPKKAK-----VLEKKSIEEKL-----NKKKEQ 4351
Qy 1895 KAAVEBQAQKRLQOQPVVIATSTSTTSSTTSISPAQKVMVAPISGVTGTGMWLT 1954
Db 4352 TESAIDEKSQK-----AEVSEIVSEKITDEKAQESQKKEVKS-----4389
Qy 1955 TKVGSPTVTFQONKNFHTATWVKQGSNGVGVQOKV--LGIIPSTSTGTSQOFTTSF 2013
Db 4390 --EAKPKKAKVLEKKSIEEKLDDKKEKOTESAIDEKSQAEEVSEIVSEKITDEK-----4442
Qy 2014 QRTATVTRPNTSGGTTNSQVITGPIQIRPGWTVIRTPLOQSTLGHAIITPVMWOP 2073
Db 4443 -----AQESQKEEVKDSEAKPKKAKV---LEKKSIEEKLDDKKEKOT 4482
Qy 2074 GAPQOVMTQIIRGQPVSTAVSAPNTVSTPGQKSLTSATSTSNIOSSASQPPRPOQ----2129
Db 4483 DSAIDEKSQ---KAEVSEIVS---EKITDEKAQES-----QKEEVKDSEAKPKKAKVLEKK 4532
```

RESULT 13

Q8MLD9

ID Q8MLD9 PRELIMINARY; PR1; 9270 AA.

AC Q8MLD9;

DT 01-OCT-2002 (T-EMBLrel. 22, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE CG18255-PA.

GN STRN-MLCK OR CG8304 OR CG18255.

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

```

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleeb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnick S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernier B., Carlson J.W., Celniker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AE003808; AAM70936.1;
RA FlyBase; FBgn0013988; Strn-Mick.
RA InterPro; IPR003961; FN.III.
RA InterPro; IPR007110; Ig-like.
RA InterPro; IPR003598; Ig_c2.
RA InterPro; IPR003006; Ig_MHC.
RA InterPro; IPR003596; Ig_v.
RA InterPro; IPR000719; Prot_kinase.
RA InterPro; IPR002290; Ser_thr_kinase.
RA InterPro; IPR001245; Tyr_kinase.
RA Pfam; PF00041; fn3; 2.
RA Pfam; PF00047; ig; 29.
RA Pfam; PF00069; pkinase; 1.
RA ProDom; PD000001; Prot_kinase; 1.
RA SMART; SM00408; IGC2; 29.
RA SMART; SM00406; IGV; 4.
RA SMART; SM00220; S_YKC; 1.
RA SMART; SM00219; TYRK; 1.
RA PROSITE; PS50835; IG_LIKE; 26.
RA PROSITE; PS00290; IG_MHC; 1.
RA PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
RA PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
RA PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Immunoglobulin domain; Transferrase.
SQ SEQUENCE 9270 AA; 1036749 MW; 0BD82D040B27C2AD CRC64;
Query Match . 4.3%; Score 641; DB 5; Length 9270;
Best Local Similarity 18.4%; Pred. No. 3.le-18;
Matches 601; Conservative 538; Mismatches 1187; Indels 938; Gaps 135;
QY 5 EEEEDGDA-----ETQDSE-----DEEMEDDDDDSYPEMEDDDDDASYCT 51
DB 2568 ESQNQKDAAGDIRKSETEDVDVHSIEKKIEPKRSEKDKLDKDFLEE-----L 2618
QY 52 ESSFRSHSTYSSTGRRKPRVHRPRSPITLEK-DIPLEFPKSSSEDVMPNHNIAI 110
DB 2619 KASAKKGQDQIEQSKQPEV-----SEVAAKISEGKIEEPKKPEEM----- 2661
QY 111 YEVLNFGTVLRLSPRFEDFCAALVSOEQCTLMAEMHVLLKAVLRPE--DTNNTFTGP 168
DB 2662 -----DTEAKSEKATVLDKQVLEKELEASAERQGD 2692
QY 169 ADLKDSVNSTLYFDIGMTWPEVLRYVYCESDKEYHHVLPYQEAEDYPYGPVENKIKVLOFL 228
DB 2693 QDVEKKSKQ-----PEVSEVAAEKISE-----ETIEPKKP----- 2723
QY 229 VDQFLTNIAREELMSEGVIQYDDHCRVCHLGLDLCCTCSAVYHLEKVPKPLEEVPED 288
DB 2724 -----EVKTEIKSEKATALD-----KQVLEKELE 2749
QY 289 ---EMQCEVCVAHK--VPGVTDCVAE-----TQKNKPYIRHBPPIGVDRSRKRYWFLNR 336
DB 2750 ASAQKQCDQDVEKSKQPEVSEIVAEEKISEKTIEEPKKPEVKTEIKSEKATA---LDK 2805
QY 337 RLIIIEDETE-----NENEKKIWIYTKVOLAEILCDLCKDYWEALCKILEEMREIHR-H 391
DB 2806 QVLEKELEASAQKQGDQDVEKSKQPEVSEV-----AE--KISEETIEEPKKPE 2854
QY 392 MDITEDLTNKGSKNSFLAAANEELLESIRAKKGDID----- 429
DB 2855 VKETEVRSEKATVLDKQVL-----EKELEASAQKQGDQDVEKFKQAEVSEVAAEKISEET 2911
QY 430 -----NVKSPETEEDKNETENDSKDAFKNREEFEDQSLEKDS----- 467
DB 2912 IEPPKPEVKDTEIKSEKATALDKQVLE--EKELEASAQKQGDQDVEKSKQPEVSEVIA 2969

```



```
Db 4793 SI-----EELKEDKKEKQTESAIDSKQAEEV-SEIVSEKITDEKAQESQKKEVDSKAP 4848
Qy 2452 -----QOIKLQPIQIOQSSAV-----OPHQIONVVTVOAASVQSOQLRQVQOIRD 2496
Db 4849 KKAKVLEKSTIEEKLKKEKQTESAIDSKQAEEVSEIVSEKITDEKAQESQKKEVDSKAP 4908
Qy 2497 QOQKKQOQIOLNVTNPSKLLIKVITQKQVVMKHNVAIEHLKQK-----KSMTPAE 2548
Db 4909 SEAKPKKAKV-----LEKKSIEEKLKKEKQTESAIDSKQAEEVSEIVSEKITDEK 4962
Qy 2549 REENORMVTCNOVMK-----YILDK--IDKEKQAAKKRKRREESVEQKRK-----QN 2594
Db 4963 ARESKKEVKSEAKPKKAKVLEKKSIEEKLKKEKQTESAIDSKQAEEVSEIVSEKITDEK 5022
Qy 2595 ATKLSALLFKHKEQLRAELKRRALLDKDLOLQTOEVLKRDOLKTKKCKDLMLQAAQATAVA 2654
Db 5023 ITDEKAQESQKKEVDSKAPPKKAKILEKKSIEIEK-----LDEKKEKQ-----TETKVA 5072
Qy 2655 APCPPVTPVLPAPPAPPPPPPPVQHTGLLSTPLPVASQKRRKREKDSKSKKKK 2714
Db 5073 TDTKSQT-----VEVSEIVLEKISEEKAESQKVELKDSKSKKAK 5114
Qy 2715 MISTTSKETKDKTKLYICKTPYDESKFYIGCDRCQNNYHGRVCVILQSEAEILDEYVCP 2774
Db 5115 VLE-----KKST-----LKEKLDENKKQKEDGATN-----KSOKAEAAD--VVP 5152
Qy 2775 QCQSTEDAMTVLTPTEKDYEGKLKVLRSQAHRKMAWFFLEPVD-----PNDA 2824
Db 5153 EKISEKVAEIKTP-----EPMDSKAKSKPDGUPADEKS 5186
Qy 2825 YGVTKPEPMDLATMEPRVQRYEKLTEFVADMTKIFDNCRYNPSDSFFYQCAEVLRSF 2884
Db 5187 HGAKVSESVPVKNEAKTKDLSAKKPTVLDEDLVV-----PKRKP-YLAEQTADSI 5236
Qy 2885 FVQKLKGF-----KASRSHNNK 2901
Db 5237 SLQTKYSMDSEYKDRKRSRKA 5260

RESULT 14
Q9VC00
ID Q9VC00 PRELIMINARY; PRT; 2768 AA.
AC Q9VC00;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE CG13648 protein.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randle M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
```

```
RA Fosler C., Gabrieli A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "the genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR ENBL; AE003750; AAF56376.1; -
DR Flybase; FBgn0039257; CG13648.
DR InterPro; IPR001007; VNF_C.
DR Pfam; PF00093; vnc; 1.
DR SMART; SM00214; VWC; 4.
DR PROSITE; PS01208; VMEFC; 1.
DR PROSITE; PS01208; VMEFC; 1.
SQ SEQUENCE 2768 AA; 294032 MW; CA929A21774E4684 CRC64;

Query Match 4.2%; Score 621.5; DB 5; Length 2768;
Best Local Similarity 18.9%; Pred. No. 4.8e-18;
Matches 584; Conservative 448; Mismatches 1197; Indels 865; Gaps 133;

Qy 78 PILEEKDIP---PLEEPKSSDLMPNEHIMNVAIYELVNFCTVLRLSPFPEDFCAA 134
Db 84 PLITCPEVPDVPYHSPGTELSIPEKF---GCSIEEKYPGEAQVPSNPNKPCELCYC 140
Qy 135 LVSOEQCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLYIFDGMTWPELVRY 194
Db 141 INNQTKVMQ-----ECTLH-VDG-----CLPIY 163
Qy 195 CESDKYHHVLPYQAEADYPGPVVENKIKVLQFLVDQFLTNN----- 236
Db 164 NKGS-----CCPVRYSCDH-----ENEL-----FMDSTTTTTPRTTGFILASTWTP 209
Qy 237 -----TAREELMSEGV-IQYDDHCRVCHKL-GDLLC-----CE-----TCSAYVHL 275
Db 210 PTTTDCIHGDFIFADGASLKGKNACEHCYCMRGDIVCAVQCEVPMMAANGKSCRAMPA 269
Qy 276 ECVKPPLEEVPEDEWQCEVCVAHKVPGVTDV-----AEIQKNKPYIRHEPIGYDRSR 329
Db 270 EGECPCSNVYCEDDSSTTEIVETTPESATSVPAKGIHAEIPKEDVDLQEHDDDKNKE 329
Qy 330 KYWFLNRLIIEEDTENENKIKWYSTKVLAEIDLCDKDYW---EAECLKILERMRE 386
Db 330 TATIPSAELSGSEIEEEEEKD---KATVAPQVTD--EKDFSEFESSTAGIPSDSR- 383
Qy 387 EIHHRMDITDLTKARGSNKSFLLAAANEIILESIRAKK-----GIDNVKSPETEKKDN 442
Db 384 -----IDLPSSTEESKESSTEAABEDIVKIVTTPPEGSGEEDVPKPSQIPEKIT 434
Qy 443 ETE-----NDSKDAENREE-----FEDQ-----SLEKSDDKTTPDDDDPOG 479
Db 435 EDELKIVTSAPAKASPEEVEVVTATSPATEEDVKPTTAGTISEEEBCKFTP---AEEG 491
Qy 480 KSEE-----PTEVGDGKNSVSANLGDNTNATSEETSPSEGRS-----P 518
Db 492 SCEEEKVKVYTAAPETEDAKPTSAV---ASDEKEQEPKPSGSGDELDLKTPTAPT 548
Qy 519 VGCLSETDSSNMAEKVASELPQDVPE-EPNKTCESNTSATTTISIOPLNENSSSEL 577
Db 519 VGCLSETDSSNMAEKVASELPQDVPE-EPNKTCESNTSATTTISIOPLNENSSSEL 577
```


Db 549 AGATSASESEBODECK -STEAPTSYDDIEPAKPTESSEASG-----EGEDVAKET 599
QY 578 NSSQSESAADDPENGERSHTPVSI-QBEIY-GDFTSEKSTGELSESPGACKGASGT 635
Db 600 TPAGEASIAAGEEIVKGTTPAGEPSEGDEEIVKGTTPAESSESEDE----- 648
QY 636 RIITRLRNPDSKLSQLSKSOVAAAHEANKLKEGKE--VLVNSOGEISRLSTKKEVIM 693
Db 649 --LTKVTTTPAGESVAGEEIEAKETTPAGEPSIAGEEIVKVTTPAGE--SSIAGEEIVK 705
QY 694 KGINNYFKLQBGKRYVHNOYSTNSFALNKHQHRHDHKKRHLAHKCLPAGFKWN 753
Db 706 -----VTTPAGESEGEIEIK-----VTTPAGESSSE 734
QY 754 GSVHGSKVIT-----ISTLRLITOLENNIPSGFFHPNMAWRANIKA--VOMCSKPRE 806
Db 735 GDBEIVKESTPAGEPISGEEDVIAKATTSAPKSDIEGVKEPETATEVPAEEVEDFAKP-- 792
QY 807 FALALAILECAVKPVVWLPITWREFLGHTRLHRMTSIEREKEKVKKKKEQBEETMQOA 866
Db 793 ---TTPIAEABEPIAGTPIPTDGIS-----GEEIVKGTTPOTLEEQPEISEES 839
QY 867 TWKYTFPVYKHQWKQGEYRTGYGGWISWKTHVYFVPVKLPNGTNNV-YRKSLEGT 925
Db 840 TEV-----PVAED-----DLSSTASASIASSTEGV 865
QY 926 KNNMDENMDESKRKSRSPPKIKIEPDSEK-----EVRGSDAAKADQN 971
Db 866 QDAASETTSSAPARAGKDEAATTVPTAQDKDDEVEQDATDLPVEDVQVSTAKTITTE 925
QY 972 EMDISKITEKKODVKELSDSDCKEPEEVEDDDMK----- 1010
Db 926 QPKEESTAEAEAEIEVTTSSPADKOEVPPEAPADKHDEEDVQATDLPKISDIGPPV 985
QY 1011 --TESHVNCQESSOVWVNVSEGFHL-RTSYKKTTSKSLDGLLERRIKOFTL-----E 1061
Db 986 VDEATNGOPEISDEATDKPSPVLPVPSQEVPSSTAKVNDNRNDFETEKTPLPSSGEDQ 1045
QY 1062 EKORLEKIKLEGGIK-----IGKTSNSSKNLSESPVITKAKEQCSDSMRQESNA 1115
Db 1046 SSEPLPAMDLPAGIPGEGCLVEGKTYANNTIVPATAPCDVCK--CISLVACQOMECK 1103
QY 1116 NNDQPE-----DLIOGCSQS---DSSVLRMSDPSHTTNKLPKDRVLDVDSVRSPTKC 1166
Db 1104 LPLENLEKCTVAADLLDGCCTPYICDESTESAKEDEESTAK--PDNKIDEDVSEISTE-BI 1160
QY 1167 PKONSTENDIEEK-----VSDLASRGOEPTKSKTKGNDFFIDDDSKLASADDI 1213
Db 1161 PRDVIPTGITEQPLSHVXPDBEIQPVTVSQAQFDESTAKVDKKP--IDES----- 1210
QY 1214 GTLICKNKKPLIOEESDTIVSSSKSALHSVPKSTNDRDATPLSRAMDPEGLKGCDSSEN 1273
Db 1211 ----AEDKKPIGESEED-----SKPIDESEEDKKPVE----- 1238
QY 1274 STLENSDRVTSQDSSEDMIV-----QNNESISGEFRFREQD-----VEVLEPLKCE 1322
Db 1239 ---ESAEDKKPVEDSESEKPLFTVIPASEIEKESKEPEDEKTEADFAAPTEQPEATTPA 1295
QY 1323 LVSSEGTGNCEDRLPVKGTANGK---KPSQOKKLEERPVNKCSDQI---KLKNTTDDKN 1376
Db 1296 QIADTAEKVEDDKLATTSPVSGEDELKPADKKRTE-----TAQIPDAEIPASTDEPE 1349
QY 1377 NENR-----ESEKKQRTST--FQINGKONKPKIYLKGECLKEISESRVSVGNVPKVNNI 1430
Db 1350 SSTELPTVLDKKPREDSKTGTEAPESDKVPEVPTSASTENEIEESDKETTIVAPPKISAS 1409
QY 1431 NKLIIP--ENDIKSLFVKE-----SAIRPFTNGDVI-----MEDFNERNSSETKSHL 1474
Db 1410 DETEPTAEEDLVPATFPIESEFEVSTKKPVGQGPLPTLAPAQPEKKPVDAETSTEADI 1469
QY 1475 LSSSDAEGNVRDSLETPSTKSDSDSTQTTTP--SACPESNSNOVEDMEIETSEVKKKVS 1533
Db 1470 STEPSAEVEKEASGETSSEDNEDAGASSTPVPVSADEKDKTPSTKTVE---ADDKFTTV 1526

QY 1534 SPIT-SEESN---LSNDFIDENGLPI-----NKNENVNGESK-----RKT 1570
Db 1527 APLAGDEESNLPKLPQDIFEEEA-PVAVTTAAPSDD---GEQKPVVEEKEPIEDGQKP 1582
QY 1571 VITEVTTMTSTVATESKTVIKVEKGDKQIOWVSTENCAKSTVTTTTTTVKLTSPSTGGS 1630
Db 1583 IEDETSTPTS-----SENEIEPESDRATTIAPSKEE--PSEPSTGAPTDEPAEPSTADP 1635
QY 1631 VDIISVKEOSKTVVTTVTVDLSLTTTGGTLVTMTSVSKYSTDKVKLMKFSRPPKRRSGT 1690
Db 1636 ESDS-KEPSESEVPTTAPA---GEKPTSSITDEEPTATSPAKPDEDEVEKETST 1690
QY 1691 ALPSYKRYTKSTKSFIVLPNDLKLARKGIGREVPYFNYNAKPAIDIMWPSPRPTF 1750
Db 1691 EIPTDAPASSEEDENS---STDQIPS-----EVP-----EKKP----- 1720
QY 1751 GITWRYLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGG---GSTRTETSETETITTE 1807
Db 1721 -----ETPAQTPEEGDIVGATAAPTTSDEVPVQ 1749
QY 1808 IIKRRDVGPIGRFYCIRKIIICPIGVPTPKETPTPQRK-GLURSALRPKRPTKQGT 1866
Db 1750 RLPE-----EVLAEIPOPSTETGKQODETTAAPSIDRK-- 1783
QY 1867 PVIIETWABEELELWEIRAFARVEK--EKAQAVEQ-----QAKKRLQEQKPTVIATST 1919
Db 1784 ----EPYVEIDEEATTVAIPSEKDEKPTEEKPVQKPTGEBPEEEKEKPIEDOVST 1839
QY 1920 TSPSTSTSTISPAQKVMVAPISGSVTTGTKNVLTTKVGSPTVTTFQONKNFHQTATWV 1979
Db 1840 EGPVSTEAASEAGSTES-----SEEVKSTPEGEVAEKPED----- 1873
QY 1980 KOGOSNS-GVVOVQOQVGLIIPSTG---TSQOTFTSFOPRTATVIR-PNTSGSGGTTTS 2034
Db 1874 KQSSSTAQAPVETIPIESTELPAQDGDKPTSEAPVSDSDTSPSDEKIPSVSG----- 1927
QY 2035 NSQVITGPQIRPGMTVIRTPLQOSTLGKAIIRTPVMVQGAPOQV-MTQIIRQOPVSTAV 2093
Db 1928 --EEVEGPEV-----ITASQAABEDELKTPAESPSSTDKVPETEYQKPEDETKAD 1977
QY 2094 SAPNTVSSTPGQKSLTSATST-----SNTQSSASQPPRQOQOVKLTMAQLTOLTQGHGG 2148
Db 1978 ETPESTVQSDVATSTSAVAGDIEKDEQATTASP--EEEEIKPTIAPAAEIPQ---- 2031
QY 2149 NQGLTVVIOGOGTTGLOLIPQGVTVLP--GPGQOLMO-AAMPNGTVOREFLETPLATTA 2205
Db 2032 -----PSEKEPVDE-QEVESGTATPAESDQPIDETATPATSPIDE-----ASTA 2076
QY 2206 TTASTTTTSTTAAGTGBQROSKLSPQMVOHQDK-----TLPPAQSSSVGPAKAQPOQT 2259
Db 2077 APTKEESTTVASAA-----SP--AVHDEIKDVTTTQPVADKEVAAPQDETKT 2123
QY 2260 A-----QPSAR--PQPTQPOPSPAQPEVQTOPEVQT---QTTVSSHVPS-----EAQPTH- 2304
Db 2124 SIDVSTDSQAQDDEKQKTEAPVAPTIVSSPTADSAASTSTVEVPSPVEIDTKPMD 2183
QY 2305 --AQSCKPQA-AQSQPQSNVQSGSPRV--QSPSOTRIRPS-----TPSOLSPQOQS 2353
Db 2184 IMSOTTAPHTADGAASTSTDEDEDAQPVTVSPQDAEKTVPSPAPQDSDKTSPSEAPODADE 2243
QY 2354 VOTTTTSQ-----PIPIOPHTSLQIPSGOQSPQVQSSTQTL-----SSQTLNQVS 2401
Db 2244 IPATAPFLDNKNIPATVAPQTDGVPATAAPLDEDKIQTATAAPLDEEKIPSTAAPLDDEK 2303
QY 2402 VSSPSRQLOIQOPQPOVIAVQLOQOVQVLSQIQSOVVAQIQAOQSGVQPKLQLPQ 2461
Db 2304 IPAPVSPVFDVBPSEKPAVSEYDGE-----ESTEPVHDVETSTDEPTSDAKLAPPTS 2358
QY 2462 IQGSSAVQTHQIONVVTVQAA-----SVOEQLQRVOOLRQOQKKQKQOIEINVTNPSKL 2516
Db 2359 APATPSESATAEIVPETAAPAELEKEVPEKATEQPELEKETPEKATEQPELEKETPEKA 2418

```
QY 2517 LIKVEI-----IQOVVVKHNAVIEHLKOKKSWTPAER-----EENORMIV 2557
DB 2419 TEQPELEKETPEKATEQPELEKEVTDKATEQSPESVDEKTTPEPVVVKSLDSTEED----- 2474
QY 2558 CNQVMKYITLDRKKEEQAOKKRRESVEQKRKQKQKATLKSAL----- 2601
DB 2475 -----ESVESEESADKKDKKTEEDTDKKHEPEVPVAVVSEIPOPSEEAAPTGT 2525
QY 2602 --LFFKHQDLRAELIKKRALDKDLQTEVQBELRDLKIKKEKDLQMLAATAVAAACPP 2659
DB 2526 HPLFPH-----LASSTTPPAVDDR-----VGEE-----DEENTVTKLSSTTTTSTTESP 2570
QY 2660 VT-----PVLPAAPPSPSPPPGQVHTGLLSTPTPLVASOKRKRREKDSKSSKKKKM 2715
DB 2571 VTSAPSTTTVASQQOQPTTPPYG--HA-----PEYEDYDEEEVFGPCTCRYAGKL 2620
QY 2716 ISTTSKTKDKTKLYCKICKTPYDESKFYIGDCRQNVYHRCVGILOS-----BAEL 2767
DB 2621 VYSAQIIPRDDPCDFCF-----FRSDIICLQQSCPPPIAGCHEEP 2661
QY 2768 IDEYVCPOCQ-----STEDAMTVLTP 2788
DB 2662 ISGFCPCRYECPVSMVAALNTTSTTTSTTLPP 2695

RESULT 15
Q9NJ17 PRELIMINARY; PRT; 5476 AA.
AC Q9NJ17;
AT 01-OCT-2000 (TREMBLrel. 15, Created)
CT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Split ends (CG18497-PC).
EN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20157049; PubMed=10655223;
RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Laverty T.,
RA Suh C., Vos M., Williams A., Rubin G.M.;
RT "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated
RT Protein Kinase Signaling Pathway That Interact With the yan Gene of
RT Drosophila Identifies split ends, a New RNA Recognition Motif-
RT Containing Protein.";
RL Genetics 154:695-712(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Boutch J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foslter C., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glisler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
```

```
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese H.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Ye J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez C., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragass V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M.B., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF184612; AAF26299.1; -.
DR EMBL; AE003590; AAN10511.1; -.
DR HSSP; P09851; 1HA1
DR FlyBase; FBgn0016977; spen.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS0102; RRM; 3.
SQ SEQUENCE 5476 AA; 590531 MW; 93FAA8C7860770C2 CRC64;

Query Match
Best Local Similarity 4.1%; Score 613.5; DB 5; Length 5476;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;

QY 305 DCVAEIQNKPIRHEPIGYDSRKYWFLNR-----RLIIEDT 344
DB 1855 TADRAEKQ-----RHEKKEKQREKQREKQREKQREKQREKQREKQREK 1911
QY 345 ENENEKKIWIYSTKVQLAEILDCLDKDYWEALFKILEEMREITHRMDITE-DUTNKA 403
DB 1855 TADRAEKQ-----RHEKKEKQREKQREKQREKQREKQREKQREK 1911
QY 345 ENENEKKIWIYSTKVQLAEILDCLDKDYWEALFKILEEMREITHRMDITE-DUTNKA 403
DB 1855 TADRAEKQ-----RHEKKEKQREKQREKQREKQREKQREKQREK 1911
```

Db	1912	EREREKKAQDRK-----KERERELRE--KEQDRKEQKEIREKDLRKEQ	1958
Qy	404	GSNKSFLAAANEILSIRAKKGDIDNVKSPETEEDKNETENDSKDAEKNREFFEQSL	463
Db	1959	-----RERONREKRLDKDLREKEMREKREKELHREKQOREHR	2000
Qy	464	EKSDDKTDPDDDE--QGKSEPE-----VDKGNVSANLGDNTWTNATSEETSPS	513
Db	2001	EKQSRAMDVQEGRGMRLESLSYQKSKMDIAGEASSLTAIDCQHKNENAMDTIAQGT	2050
Qy	514	EGRSPVCLSETPDSSNMAEKKVASELPQDYPEENK--TCESNNTSA-----TTTISI	564
Db	2061	PGASP-----STP--SDNTPKRSRKLKNSVRUHLKRLRSQBSNHSAGGGSCGGSSH	2113
Qy	565	QPNLENSSELNSQSSESAAADDPENGERS--HTPVSIQOEIVGDFTSKSTGELS	622
Db	2114	QIHEDYVVRIMWNSQNTSVHSSNQRLNDRDRSKEHSSFKED-----KNSSSHIS	2166
Qy	623	ESPGAGKAGSGSTRIITRLRNPDKSLQKLSQOAAAAAHEANKLPKEGKEYLVVNSQGEI	682
Db	2167	RPHCGCGSSASSKHHR-----ROKHQKGSSASITNSSI-----EYVVDPIQOTK	2214
Qy	683	SRLSSTKVEIMKGINNYFKLQEGKYRYVHNOYSTNSFALNK-----HQHR-----	729
Db	2215	HNLNTSEELQSHQ-----PKREKEREHFSSHANSSSRHKSKRDHHRHKKRHSVA	2267
Qy	730	-----EDHDKRHLAHKFLT-----PAGEFKWNGSVHGSKVLITSLITLRLITOLENNIP	779
Db	2268	ESTNTDEHTPOQHNPHRISAGSGSAGELSSAATNTSSKLLHHQHHRHRSVERKSSRGS	2327
Qy	780	SSFFHPNASHRANWIKAYOMCSKPREFALALAILCAVAKPVMLPIWREFLGHTRLHRM	839
Db	2328	DEGHHSSSLRAKLMMLSSADSDTDDA-----SKKHSI	2362
Qy	840	TSITERE-----EKEVKKKEKQEBEETWQATWVKYTFPVKHQWKQGEERYVTGYG	894
Db	2363	FDIPDDCPNVMDYKARSKNMQRQAEKKIKAKFS-----QLKOSRAKKKSTSYDG	2417
Qy	895	WS-----WISKTHVYRVPKLPNGTNYNRYKSLGTKNNMDENMDSKRSRSPKKIKIE	951
Db	2418	DSDTEDFROH-----RNSGSSSFHGR--YPGLSSDDDDDEETHORRIS-----	2460
Qy	952	PDSEKDEVKSDAAGADQNE-----MDISKITEKKQDVREKLLDSDDKPCKEEPMYD	1006
Db	2461	-----SDSDAEHGGQDNQOGASTLADANRVROMQ-QNLRLCDGDDS--SED--EIR	2506
Qy	1007	DMKTESHVNCOESSQDVVNVSEG-----FHLRTSY-----KKTKSKSLDGLL	1051
Db	2507	RNYMKHSHFGKKNSTRISADSESQSPADPLTIKQEHPIAPAQEIKREOLSEDEQKFK	2566
Qy	1052	ERRIKOFTLEEKORLEKIKLEGKIGIKGTSTNSKNLSSESPVITKAKEGQCSDSMRQEQ	1111
Db	2567	SRHDSNSSIEER-----KLATEREI-----KTELGDYNSSEYTYGCKLKEYSPETRKKHK	2618
Qy	1112	SPANANDQPEDLIQCSQSDSS-----VLRMSDPS-----HTTNKLYPKDRVLDDVSI--	1159
Db	2619	SKRR-----LKSSSTADTSAAQTPLVMTPLTPIEDVHSSSECKTKFDNFDDLKTEC	2670
Qy	1160	-----RSPETKCPKQNSIENDIEKVSJDLASROEPTKSKTKGNDFIDDSK	1206
Db	2671	SSIPLEISAGERKKHREKREKRLNMTATVPN-----SPTTNDT-----SSE	2716
Qy	1207	LASADDIGTLCKNKPLIQEESDITVSSSKALHSSVPKSTNDRDAPTLSR-----	1258
Db	2717	KLSEERHRLKSKSKSDMNSCNTKIYNSSGA-H---PSTSPSLPATPTSPAPTATQSK	2772
Qy	1259	-----AMDFEGLKGCSESNTLENSSDTVSIQDSSEEDMIVQNSNESISEOFRTEQDVE	1314
Db	2773	RGEDKMEFIFGIISDEESQFFQEAETNKDIIIPSS-----VSTGTPVISAALQYKQ---	2824
Qy	1315	VLBPCLKELVSGESTGNCEDRLPVKGTBANGK-----KPSQOKKLEERPVNKCSDOI	1370
Db	2825	-----EP-----STPS-----KNEAHIOILTVEHEPOOOLER---SRLSGCSSSS	2864

Qy	2256	QPQT-AQPSARPQQTQPSQAPEVQTVQEVQTVQTTVSSHVPSEAQPTHAQSSKPOVAA	2314
Db	3760	QPQVIKMTAHQHQHMQQFHMQMIQHQHMQ-----QQLHQSQ--QITS	3805
Qy	2315	QSPQSNVQSGSPVRVQSPQTRIRPSTPQSLSPGO-QSOVQVTTTSQPIPIQPHTSQIP	2373
Db	3806	APQHMHQHQHQAQQQQHQQHQQHQQHQAQQHTQKHQAQQQFNQ--QIQHQSQQ--	3861
Qy	2374	SQGQPSQPPQVS---STQTLSSQTLNQVSVSPSRPQLIQOPQ-QVIAVPQLQQQ-	2428
Db	3862	QHQQVQONQAQQHLSQQHQSQOQLNQ---QHQAQQOQLQOIQKLOQMHHGPPQQQKS	3916
Qy	2429	-----VQVLSQIOSQVVAQIAQOQSGVPPQIKLQLPQIQOQSSAVQTHQIONV	2476
Db	3917	PQGVGHLGGSTSI PASQOHNSQLPAR-----GVPOQ---QHPQQLSHSSPCPKPNTLVSV	3967
Qy	2477	VTVQAASVQEOQLQVQQLRDQOQKKQO-----QIEINV--NTP	2513
Db	3968	NQGVQPPAILTRVGGSHSQPNQOQQLPHQQSSSGHPHOKQLSSPGANLPLOTPLNVIONTP	4027
Qy	2514	SKLLIKVEII-QKQV---VMKHNAVIEHLKQKKSMTPAEREENQRMIVCNQVMKYILDKI	2569
Db	4028	KIIVQQHIVAQNQVPPQTQGNAIHYPNQNGKDSPPGH-----	4066
Qy	2570	DKEEQAKKRRKREESVEQKRSQKATKLSAL-----LFKHKEQLRAELKKRALLDK	2622
Db	4067	VEPTPAMSAQKTSSESVIRTPPTTGLAVISANTVGSLLITEENLIKISQPKODELIEQ	4125
Qy	2623	DLQIEVQ-----EELKRDLLKIK-----EKDLMQLAQATAVAAPCPPVPLPAPPAP	2670
Db	4126	DSK-EVDSDYWSAKEVWIDSVIRKLDTPPLASKDAKRAVEMQAI-----AP	4169
Qy	2671	PPSPPPPPGVQHTGLSTPTLPVASQKRRKEEKDSSSKKKKM	2715
Db	4170	APIPNPQGNQ--SMAQETALPTTSMVNSNDHDTETETETROL	4212

Search completed: September 24, 2003, 01:16:36
Job time : 106.506 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2003, 19:30:01 ; Search time 19.932 Seconds
(without alignments)
6858.664 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

Sequence: 1 MVSEEEEDGDAETQDSE.....KLGFASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	4082	27.3	810	1	FALZ_HUMAN
2	476	3.2	3726	1	TRX_DROME
3	451	3.0	5085	1	PCLO_RAT
4	447.5	3.0	1878	1	BA2A_HUMAN
5	441.5	2.9	5120	1	PCLO_CHICK
6	439.5	2.9	3256	1	K167_HUMAN
7	439.5	2.9	3924	1	ANK2_HUMAN
8	438.5	2.9	5147	1	PCLO_HUMAN
9	422.5	2.8	1850	1	BA2A_MOUSE
10	412	2.8	5038	1	PCLO_MOUSE
11	407.5	2.7	3828	1	TRX_DROVI
12	406	2.7	2492	1	ATRX_HUMAN
13	406	2.7	4377	1	ANK3_HUMAN
14	401.5	2.7	1367	1	AMYH_YEAST
15	391	2.6	5179	1	MUC2_HUMAN
16	391	2.6	6632	1	UN89_CAEEL
17	387.5	2.6	3911	1	AKA9_HUMAN
18	377	2.5	2130	1	BA2B_CHICK
19	375.5	2.5	3259	1	GIAN_HUMAN
20	374.5	2.5	2116	1	MVS2_DICDI
21	373	2.5	2476	1	ATRX_MOUSE
22	371	2.5	1589	1	PHB_DROME
23	370	2.5	1972	1	BA2B_HUMAN
24	367	2.5	3969	1	HRX_HUMAN
25	365	2.4	2453	1	NCR1_MOUSE
26	364	2.4	2349	1	TPR_HUMAN
27	362.5	2.4	2843	1	APC_HUMAN
28	358	2.4	2464	1	NAPB_MOUSE
29	356.5	2.4	3866	1	HRX_MOUSE
30	354.5	2.4	2035	1	HFC1_HUMAN
31	353.5	2.4	3562	1	PGCV_CHICK
32	353.5	2.4	5703	1	MU5B_HUMAN
33	353	2.4	5171	1	BPEA_HUMAN

RESULT 1				
ID	FALZ_HUMAN	STANDARD;	PRT;	810 AA.
AC	Q12830;			
DT	15-JUL-1999	(Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Fetal Alzheimer antigen (Fetal Alz-50-reactive clone 1).			
GN	FALZ OR FAC1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=95347245; PubMed=7621746;			
RA	Bowser R., Giambrone A., Davies P.;			
RT	"FAC1, a novel gene identified with the monoclonal antibody Alz50,			
RT	is developmentally regulated in human brain.";			
RL	Dev. Neurosci. 17:20-37(1995).			
CC	-1- SUBCELLULAR LOCATION: DETECTED IN BOTH THE CYTOPLASM AND NUCLEUS			
CC	OF CELLS IN THE DEVELOPING CORTEX. IN THE ADULT BRAIN, IT WAS SEEN			
CC	ALMOST EXCLUSIVELY IN THE NUCLEI OF NEURONS OF THE NEOCORTEX. IN			
CC	THE BRAINS OF ALZHEIMER DISEASE PATIENTS, THE PROTEIN IS LOCALIZED			
CC	IN A SUBSET OF AMYLOID-CONTAINING PLAQUES.			
CC	-1- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THE FETAL BRAIN.			
CC	EXPRESSION IS MUCH LOWER IN ADULT BRAIN AND IS HIGHER IN			
CC	NEURODEGENERATIVE DISEASES.			
CC	-1- SIMILARITY: Contains 1 PHD-type zinc finger.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U05237; AAA97522.1; -			
DR	PIR; G01252; G01252.			
DR	TRANSFAC; T04682; -			
DR	Genew; HGNC:3581; FALZ.			
DR	MIM; 601819; -			
DR	GO; GO:0005737; C:cytoplasm; TAS.			
DR	GO; GO:0005634; C:nucleus; TAS.			
DR	GO; GO:0007399; P:neurogenesis; TAS.			
DR	InterPro; IPR004022; DDT_dom.			
DR	InterPro; IPR001965; znf_PHD.			
DR	Pfam; PF02791; DDT; 1.			
DR	Pfam; PF00628; PHD; 1.			
DR	SMART; SM00571; DDT; 1.			
DR	SMART; SM00249; PHD; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 1.			
DR	PROSITE; PS00016; ZF_PHD_2; 1.			
KW	Nuclear protein; Zinc-finger.			

P49454 homo sapien
P37929 mus musculu
P15205 rattus norv
Q04893 saccharomyc
Q02455 saccharomyc
O75376 homo sapien
P25386 saccharomyc
Q8msl1 drosophila
P51611 mesocricetu
P46821 homo sapien
Q61329 mus musculu
Q10411 schizosacch

FT	2N_FING	251	298	PHD-TYPE.	
FT	DOMAIN	4	9	POLY-GLU.	
FT	DOMAIN	30	35	POLY-ASP.	
FT	DOMAIN	42	46	POLY-ASP.	
FT	DOMAIN	657	660	POLY-ALA.	
SQ	SEQUENCE	810 AA;	91799 MW;	F7E2C992FE5BE96D CRC64;	
Query Match					27.3%; Score 4082; DB 1; Length 810;
Best Local Similarity					98.7%; Pred. No. 7e-137;
Matches 7/77; Conservative 4;					Mismatches 2; Indels 4; Gaps 1;
QY	1	MYSEEEEDGDAEETQSDDEDEDEDDDDSDYPEEMEDDDDDASCTESSFRSHST	60		
DB	1	MYSEEEEDGDAEETQSDDEDEDEDDDDSDYPEEMEDDDDDASCTESSFRSHST	60		
QY	61	YSTGGRKPRVHRPRSPILKEDIPPLEFFPKSSDLMVPHNIMNVIAIYEVLFNFGTV	120		
DB	61	YSTGGRKPRVHRPRSPILKEDIPPLEFFPKSSDLMVPHNIMNVIAIYEVLFNFGTV	120		
QY	121	LRLSPFRFDFCAALVSQOCTLMAEMHVLLKAVLREEDTNTTFFGADLKDSVNSTLY	180		
DB	121	LRLSPFRFDFCAALVSQOCTLMAEMHVLLKAVLREEDTNTTFFGADLKDSVNSTLY	180		
QY	181	FIDGMTWPEVLRVYCESDKKEYHHVLPYQAEADYPYGPVENKTKVLQFLVDQFLTTNIARE	240		
DB	181	FIDGMTWPEVLRVYCESDKKEYHHVLPYQAEADYPYGPVENKTKVLQFLVDQFLTTNIARE	240		
QY	241	ELMSEGVIOYDHCRCVCHKGLDLCCTCSAVYHLECVKPPLEEVPEDEWCQECVAVHKV	300		
DB	241	ELMSEGVIOYDHCRCVCHKGLDLCCTCSAVYHLECVKPPLEEVPEDEWCQECVAVHKV	300		
QY	301	PGVTDCAEIQNKPKYIRHEPIGYDRSRKRYWFLNRRLLIIEEDTENENEKKIWIYSTKVQ	360		
DB	301	PGVTDCAEIQNKPKYIRHEPIGYDRSRKRYWFLNRRLLIIEEDTENENEKKIWIYSTKVQ	360		
QY	361	LABELDCLDKYWEALCKILEMREIHRHMDITEDLTNKARGSNKSFLLAANEIILES	420		
DB	361	LABELDCLDKYWEALCKILEMREIHRHMDITEDLTNKARGSNKSFLLAANEIILES	420		
QY	421	IRAKKGDIDNVKSPETEKKNETENDSKDAEKNEEFQDSLEKSDSDKTPDDDPQEQK	480		
DB	421	IRAKKGDIDNVKSPETEKKNETENDSKDAEKNEEFQDSLEKSDSDKTPDDDPQEQK	480		
QY	481	SEEPTEVGKGNVSANLGDNTNATSEETSPSEGRSPVGCISETPDSSNMAEKKVASEL	540		
DB	481	SEEPTEVGKGNVSANLGDNTNATSEETSPSEGRSPVGCISETPDSSNMAEKKVASEL	540		
QY	541	PDVPEEPNKTCESSNTSATTTISIQPNLENSSSSELNSSQESAKAADDPNGERESH	600		
DB	541	PDVPEEPNKTCESSNTSATTTISIQPNLENSSSSELNSSQESAKAADDPNGERESH	600		
QY	601	PVSIQEIIVGDFTSKSTGSELSPGAGKSGASTRIITRLRNPDSKLSQKLSQQVAAAA	660		
DB	601	PVSIQEIIVGDFTSKSTGSELSPGAGKSGASTRIITRLRNPDSKLSQKLSQQVAAAA	660		
QY	661	HEANKLFKEKEVILVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKRYRVHNOYSTNS	720		
DB	661	HEANKLFKEKEVILVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKRYRVHNOYSTNS	720		
QY	721	FALNKQHRDHDKRRHLAKFCLTPAGEFKWNGSVHGSKVLTITSLRITITQLENNIPS	780		
DB	721	FALNKQHRDHDKRRHLAKFCLTPAGEFKWNGSVHGSKVLTITSLRITITQLENNIPS	780		
QY	781	SFFHPNWN 787			
DB	777	TSLHPSF 783			

RESULT 2
TRX_DROME
ID TRX_DROME
AC P20659; Q27255; Q27327;
DT 01-FEB-1991 (Rel. 17, Created)

16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Trithorax protein.
TRX.
Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=9019757; PubMed=2107543;
RA Mazo A.M., Huang D.-H., Mozer B.A., Dawid I.B.;
RT "The trithorax gene, a trans-acting regulator of the bithorax complex
in Drosophila, encodes a protein with zinc-binding domains.";
Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
[2]
RN SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RX MEDLINE=95009521; PubMed=7924996;
RA Sedkov Y., Tillib S., Mizrokhi L., Mazo A.;
RT "The bithorax complex is regulated by trithorax earlier during
Drosophila embryogenesis than is the Antennapedia complex, correlating
with a bithorax-like expression pattern of distinct early trithorax
transcripts.";
RL Development 120:1907-1917(1994).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=Oregon-R;
RX MEDLINE=96100387; PubMed=8555104;
RA Tillib S., Sedkov Y., Mizrokhi L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
Mech. Dev. 53:113-122(1995).
[4]
RN CHARACTERIZATION.
RX MEDLINE=95047388; PubMed=7958911;
RA Kuzin B., Tillib S., Sedkov Y., Mizrokhi L., Mazo A.;
RT "The Drosophila trithorax gene encodes a chromosomal protein and
directly regulates the region-specific homeotic gene fork head.";
Genes Dev. 8:2478-2490(1994).
CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
CC IsoId=P20659-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P20659-2; Sequence=VSP_006665;
CC -!- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
ZINC.
CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31617; AAA29025.1; -
CC EMBL; Z50152; CAA90514.1; -
CC EMBL; Z50152; CAA90513.1; -
CC EMBL; Z31725; CAA83516.1; -
CC EMBL; Z31725; CAA83515.1; -
CC FIR; A35085; A35085.
CC HSSP; P20393; IA6Y.
CC TRANSFAC; T00850; -.


```
DR FlyBase; FBgn0003862; trx.
DR InterPro: IPR003889; Fyrich_C.
DR InterPro: IPR003888; Fyrich_N.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001965; znf_PHD.
DR Pfam; PF00628; PHD; 3.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRIC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00249; PHD; 5.
DR SMART; SM00508; PostSET; 2.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS00868; POST_SET; 1.
DR PROSITE; PS0280; SET; 1.
DR PROSITE; PS01359; ZF_PHD_1; 4.
DR PROSITE; PS00016; ZF_PHD_2; 3.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator;
KW Alternative splicing.
FT ZN_FING 1266 1347 PHD-TYPE 1.
FT ZN_FING 1348 1393 PHD-TYPE 2.
FT ZN_FING 1421 1482 PHD-TYPE 3.
FT ZN_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
FT ZN_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3587 3708 SET.
FT DOMAIN 3710 3726 POST-SET.
FT DOMAIN 512 516 POLY-SER.
FT DOMAIN 565 570 POLY-ASP.
FT DOMAIN 661 664 POLY-SER.
FT DOMAIN 905 910 POLY-SER.
FT DOMAIN 1576 1582 POLY-GLN.
FT DOMAIN 2298 3027 GLN-RICH.
FT DOMAIN 3032 3040 POLY-SER.
FT DOMAIN 3181 3184 POLY-GLN.
FT DOMAIN 3220 3225 POLY-GLU.
FT VARSPLIC 1 368 Missing (in isoform Short).
FT /FTID=VSP_006665.
FT P -> PWLSPFLKFLGLSTHGGLLLWLLGLVGVVRLKQGG
FT (IN REF. 1).
FT R -> S (IN REF. 1).
FT CONFLICT 2341 2341 R -> S (IN REF. 1).
FT CONFLICT 2392 2392 G -> S (IN REF. 1).
FT CONFLICT 3726 3726 AA; 400575 MW; D2756E50763DLCF5 CRC64;
SQ SEQUENCE 3726 AA; 400575 MW; D2756E50763DLCF5 CRC64;

Query Match 3.2%; Score 476; DB 1; Length 3726;
Best Local Similarity 18.8%; Pred. No. 1.9e-09;
Matches 628; Conservative 419; Mismatches 1138; Indels 1154; Gaps 155;

QY 5 EEEEEGDAAETQDSDD-----EEDMEEDDDSDYPEEMEDDDDDASCTESSFRSHST 60
DB DEDEGGVTFRNDSPEDQNNAEDEDDDDDEAEEDDQNEEDDNDDEAAESEKSAETKS 597
QY 61 YSS-----TPGR--RKPRVHRPSRPILEEKDIPPLEFPKSSIDLMPN 101
DB AGADERDPEKQLVMDSHFVLPKRSTRSRRIKPNKRILLEGA-----STKKPLSLGD 651
QY 102 EHTMNVIAI-----YEYLRNFGT-----VYLRSPFRFE 129
DB SKGKNVFGTSSSSAGSTASTFSASTNLKLGKETFNFNGTLKPNSSAAGIFVLRQPRLQFQ 711
QY 130 DECAALVSOEQTLMAEHVVLLKAVLREEDTSNTFGPADLKDSVNSTLYLFDGWTWPE 189
DB ADNQATFAAPKACPTSPSAIPKPNASLATSSFGSLASTNSSTV-----TPTPS 760
QY 190 VLRVYCE---SKEYHHVLPYAEADYPYGPVENKI--KVLQFLVDQFLTNTIAREELMS 244
DB ACSI-CSAVVSKEV-----TQARKYGVACDVCRRFFKFKMTKKSISANSSTAN-TS 810
QY 245 EGVYQYDDHRCVCHKLGLDLLC-----CETCSAVYHLECVKPPPEVEPEDEWQCEVC-- 295
DB SSSQY-----LQCKNGEGSPCSIHSAKSQLNKFKFYKD-----RCTACWL 852
QY 296 -----VAHK-----VPGVT-DCVAEIQKNKPYIRHEPIGYDRSRKRYWFLNRR 337
```

```
853 KKCMISFQLPAAHRSRLSAILPGMRGEAAAREKSAELL--SPTGSLRF----- 900
338 LIIEDTENENKKIWIYSTKVLQELADCLDKDYWEA--ELCKILIEEREEIHRHMDIT 395
901 ----TSTASSSPSV-VASTSVK-----WKSGDSTSALTSIKPNPLAENNVIT 943
396 EDLTNKARGS---NKSFL-----AAANEELIESIRAK--KGDIDNVKSPETEKKD 441
944 FGSTPLLRPAILENPLFLKISNAADQLAAAEAISPLTKKNSKQEKVKESQSEKLL 1003
442 NETENDSKD---AEKNREFEQSLKSDDDKTPDDDDPEQKSE-----EPTVEGD 489
1004 SPTQAGTKKSGAAEAQVEEVQKEEAPQTSITTPTQPSASNGASHGVPQAEAGETNATGD 1063
490 KGNVSANLGDNTTATSEETSPSEQRSPVGCLE--TPDSSNMAEKKVASELPDQV--- 544
1064 TLKQRIDILPGPRVKHVCRSASIVLQG-PLATFGEQOQDEADMQEETAAAPVPSAIMEP 1122
545 -PEPNKTCESSENTSATTTSIQPNLENSNSSELSSQSESAKAADDPENGERSHPTVS 603
1123 SPEKPHIVTDENDNCASCTSPVGDESKPSKSSGSAQAQAEVKA-----TA 1168
604 IQEEIVGDFTSEKSTGELSESPGAGKAGSGSTRIITRLRNPDSKLSQLKSQQVAAAHEA 663
1169 LGKE-----GTASAAGSSAKVT-TRN-----AASVAS 1194
664 NKLFBEGKEVLVNVNSOGELSRSLTKKEKIMGNINNYFKLGOEGKRYRVHNOYSTNSFAL 723
1195 NLIVAASKK---QRNGDI---ATSSSVTQSSN-----QTQGRKTEKHRQRT----L 1236
724 NKQHREDHD-----KRRHLA-----HKFCLT 745
1237 ISIDFENYDPAEVCOTGFGLIVTETVAQRALCFLCGSTGLDPLIFCAACCEPYHQYCVQ 1296
746 PAGEFKWNGSVHGS---KVLTTSTLTLTTOLENNIPSSFFFPNASHANWI----- 795
1297 DEYNLK-----HGSEFDTTLMGSLLETTVN--ASTGPFSSSL--NQLTQRLNMLCPRCTVC 1347
796 -----KAVQMCSEK-----PREFALALALECAVPPVVMPLPIW 827
1348 YTCNMSSGSKVKCQKQKNYHSTCLGTSKRLLGADRPCLVCNCLCKCKSSTTKV----- 1401
828 REFLG-----HTRL----- 836
1402 SKFVGNLPMCTGCFKLKRGKGNFCPCQRCYDDNDFDKMMEGDCQGWHSKCEGLSDEQ 1461
837 HRMTSIEREEKEKVKKKKQEEETMQOATWVKYTFPVKHQVWKQK-GEEYRVTCYGGW 895
1462 YNLLSTLPESIETICKKCARNESS-----KIAEEMWQAVMEEFKASLYSVL 1509
896 SWTSKTHVYRVFKLPNGTNNVNYRKSLEGTKNNDENMDSEDKRCSRSPKKIKIEP--- 952
1510 KLLSKRQACALLKLSPRKNV-----RCTCGASSNOGKLOPKAL 1548
953 --DSEKDEYKSDAAKADQONEMDISKITEKDDQVYKELLDSDSKPCKEPEMEVDDMK 1010
1549 QFSGSDNGLGSD---GESONSDVYEFKDDQOQ--QOQRANMMNKP-RVKPL-----PCS 1598
1011 TSHVNCQRS-SQVDVNVNVEGFLHRTSYKKTKSKKLLGDLLERRRKTQBLEEKQBLEKI 1069
1599 COOHTSHSQSFLVDI-----KOKIAGNSYVSLEEFYDMYSQVIOQ- 1639
1070 KLEGGIKIGKTSNSSKNLSESPVITKAQEGQSDSMRQESPNANNDQ-----PED 1122
1640 -----SNCDELDIAYK-----ELLSEQFPWFQNETKACTDALEED 1674
1123 LIQGCC-----QSDSSVLRMSDPSHTTNKLYPKDRVLDVYSIRSPETKCPKQNSIEND 1175
1675 MFESCSGGNYEDLQDAGGVASVYNHSTSQAESRSGVLD-----IPL----- 1717
1176 IEKIVSDLASRQOEPTKSKTKGNDFIDDDSKLASADDITLICKNK-----KPLI 1225
```

```
Db 1718 --EEVDGSGGIG-MRLDTRMCLFCRKSGBGLSGEERALLYCGHDCWHVHTNCAMWSAEV 1774
QY 1226 QESDITVSSSKSLHSSVPKS-----TNRDATPLSRAMDFEGKLGCDSES-----N 1273
Db 1775 FEED-----GSLQVHSAVAGRMKCTVCGNRAT-----VGCNVRSCGHHY 1819
QY 1274 STLENSDTSIODSS-----EEDMIVQNSNESISEQFRFTREODVEVLEPLKCLVS 1325
Db 1820 YPCARSIDCAFLTDKSMYCPAHAKNGNALKANGSPSV-----TVESNEFSRPPYVEL-- 1872
QY 1326 GESTGNCEDRLPVGTGANGKPSQOK-----KLEER-----PVN-KCS 1363
Db 1873 -----DRRKKLLIE-----PARVQFHIGSLVRQLGAIVPRFSDSYEAVVPINFLCS 1919
QY 1364 -----DOIKLKNTDKKNENRESE-----KKQRTSTFOINGKDNKPKIYLKGECLKE 1412
Db 1920 RLYSSKEPWKIVEYTVRTTIONSSTLTALDVGRTVTDHTNPNSEKVLGMAIARWH 1979
QY 1413 ISESR---VVSNGVE-----PKVNNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNE 1464
Db 1980 TSLARSELENGGTDWSEGFNPNSC--VPPDQNTSEEPQOQADLLPELKDALFEDL-- 2035
QY 1465 RNSSETKSHLSSDAEN---YRDSLETLPSTKES-DSTQTTTPSASCPESSNSVNOVED 1520
Db 2036 -----PHELDDGISMLDIFLYDDKDLFAISQSKDGTQAMTS-----NQAQN 2078
QY 1521 MEIETSEVKVKTSSPITSEESNLSNDFIDENGLPI-KNKNENVANGESKRKTIVTEVTMT 1579
Db 2079 QNOQAGGANSVS-----ICDETRNSNTSLG-NGWPASNVEDAMLSAARN---SSQVQML 2130
QY 1580 STVA-----TESKTVIKVKGDKQTVVSTENCASKSTVTTTTVTVKLSTPTSTGGSDV 1632
Db 2131 KTLAWPKLDGNSAMATAIKRKLK-----NLAEGVFLLSSQQRNKKEMATVAGV- 2181
QY 1633 IISVKEQSKTVTTVTDSLTTTGGTLVTSMT--VSKEYSTRD-----KVKLMKFSRP 1683
Db 2182 -----SRQISSETSEVEGATTSVGSVRKSTFWSAAKRYFEKSEGREBAKRLQMDG- 2235
QY 1684 KTRSGTALPSYRKVTWKSTKKSIFVLPNDLKLARKGIR-----EVPYFNYNAPALD 1739
Db 2236 -----VDDSITEFRIISGDGNLSAQFSGQVKKDRCCQCTRYNYDAFOR-- 2278
QY 1740 IWPYSPRPRTGITWRYLQIVKSLAGVSLMLRLLLWASLRWDDMAAKVPPGGSTRTETS 1799
Db 2279 --HLPSCSPT-----MSSNETESDVS-----GOGMTNNATQ 2307
QY 1800 FTEITTTTELKRRDVGPIRFEXCIRKIICPIGVPEPKETPTPQR-----KGLR 1850
Db 2308 ISAESLNEQLQLLANAGLNY-----LQSATFPQVQRLGSLGQFGLQLQ 2354
QY 1851 SSALRPK-----RPETPKOTGPVILITWVAEELEW--EIRAFARV----- 1891
Db 2355 QLQLQPSGLNGFFLSQNPATQANT-----DDLQIYANSLQGLAANLGGGFTL 2403
QY 1892 -----EKEKAQAEVQQAQKLEQKQKTVIATSTTSPTSTSTISPAQKVMVAPI 1941
Db 2404 AOPTVTAPAQPOLIAVSTNPGTQOFOIQTPTMOATTTPTATYQTLQATNTDKKIML--- 2460
QY 1942 SGSVITGKMWLVT--TKVGSFATVTFQKNKPHOTFATWK-----QCGSNSGVYVQVQ- 1993
Db 2461 --PLTAAGKPLKTVATKAAQAAVQKRLKSHGVKPKLQAKLPHQPQHQQOQQTQVQOP 2518
QY 1994 -KVLG-----IIPSSTGT-----SQOFTSTF----- 2013
Db 2519 ITVMGNLLQOLLQPSSTQTOAPQIILPQAPQNIISFVTGDSGQGPLOYISIPTAGE 2578
QY 2014 ---QPR-TATVTIIRNTSSGG-----TTSNS--QVITGPQIRPGMTVIR 2052
Db 2579 YKPQOPTATPFLITAPACAGATYLOTASGNLVLTTPSNSGLQMLTAQSLQAPQVIG 2638
QY 2053 TPLQOQSTL-----GKAIITFPVMVQPCAPQVM-----TQIIRG 2086
Db 2639 TLIQPTIQLGGGADGN-----QPGSNQOPLILGGTGGSSGLEFATTSPOVILAT 2689
QY 2087 QPV-----STAVSAPNTVSSST-PG---QKSLTSATSTSIQSSASOPPRPQOGVK 2133
Db 2690 QPMYVGLGTIVQNTVMSQQQFVSTAMPGLMSQNASFSATTTQVFOASKIEP-----IVD 2743
QY 2134 LTMALQLTOLTOGGHNGOGLTV---VLOGOQ--TTGOLQLIPQGVTVLPQPGQQLMQAA 2187
Db 2744 LPAGYVVLNNTGDASSAGTFLNAASVLQOQTODDTTQI----- 2782
QY 2188 MPNGTVQRFLETLPLATTATTASTTTTSTVSTAAGTGEORQSKLSPQMQVHQDKTLPPAQS 2247
Db 2783 LQANFQ-FQSVFTSSGASTSMDYTSFVMTA-----KIPPYTIKRTNAQAKAAG 2832
QY 2248 -SSVGPAKAQO-----TAQPSARQPOTQPSQPAQPEVQTOPEVQTOFTVSSHVPS 2299
Db 2833 ISGVGVPPQPVVKNVLPPTSIVTQSQVQVKNLKSQVKGKAASGTGTTCGA-PPSI 2891
QY 2300 A-----OPTHAQSSKPOV-----AAQSQPSQNVQSQSPVRV--Q 2331
Db 2892 ASKPLQKKTNMRIRPIHKLEVKPKVMKPTPKVQNONHSLLQOQOQOQPOLQOQIPAVVNVQ 2951
QY 2332 SP-----SOTRIRPST-----PSQLSPGQSQVQTTTTSQPI-----PIQPHTSLQ 2371
Db 2952 VPKVTISQQRIPACTOQOQLQQAQMIHIPQOQOPLQOQOQVQVQSPMPTITLAEAPVVQSQ 3011
QY 2372 IPSQOQPSQPSQVOSSTQTLSSGGTFLNQVSVSPSR----- 2407
Db 3012 FVMEPQALEQOELANRVOHFSTSSSSSSNCSLPTNVVNPMMQQAQFSTTSSTTRPTNRV 3071
QY 2408 -POLQIOQPOP-----QVIAVPOLQOQVQ--VLSQIQSQVVAQIQAOQSGVPOQIKLQLP 2459
Db 3072 LPMOOREPAPLSNECPVSPSPTPPKPVEQPIIHQMTSASVSKYAKSTLPSPV---YE 3128
QY 2460 IQIQSSAVQTHQIQNVTVTVOAASVQEQ-----ORVOQLRDOQOQKK 2502
Db 3129 AELKVSSVLES--IVPDVTMDAILEEQPVQTSIYTEGLYEKNSPGESKTEQLLLQOQORE 3186
QY 2503 Q--QOIEINVTNPKLLIKVEIIOKQVVMKHNVIH-----LKOKKS-MTPABREEN 2552
Db 3187 QLNQOOLVNGYLLDKHFTQVEPMDTDVYREEDLEEEDEDDDFSLKMRTSACNDHEMSDS 3246
QY 2553 QRMIVCNQVMKYIILDKIDKE-----KQAQKKRK 2581
Db 3247 EEPVAKDKISK-ILDLNLTNDCCADSIATATTMEVDASAGYQOMVEDVLATTAQASAPTEE 3305
QY 2582 REESVEQKRSQNKATKLSALLFKHKEQLRAELKRALLD-KDLOIEVQBELKRLDKK 2640
Db 3306 FEGALETAAVEAAATYINEMADAH-----VLDLQQLONGVELELRRR---KE 3349
QY 2641 EKDLMLQAOATAVAPCPPTVPLPAPPAPPPPPPPPG 2679
Db 3350 EORTVSQEQEOSKAAIVP--TAAAPPEPQIQEPKKMTG 3386
RESULT 3
PCLO_RAT
ID PCLO_RAT STANDARD; PRT; 5085 AA.
AC Q9JKS6; Q9JUL1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN PCLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
```

RT bassoon.*;
 RL Neuron 25:203-214(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
 RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
 RP ALA-4694.
 RX MEDLINE=21181819; PubMed=11285225;
 RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
 RT "An unusual C(2)-domain in the active-zone protein piccolo:
 RL Implications for Ca(2+) regulation of neurotransmitter release.";
 EMBO J. 20:1605-1619(2001).
 CC -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin.
 CC -1- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
 CC junctions.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9UKS6-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9UKS6-2; Sequence=VSP_003930, VSP_003931;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF138789; AAF07822.2; .
 DR EMBL; AF227534; AAF63196.1; .
 DR HSSP; P04410; 1A25.
 DR GO; GO:0045202; C:synaptic junction; IDA.
 DR GO; GO:0005509; F:calcium ion binding activity; IDA.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. . . ; IDA.
 DR GO; GO:0005522; F:profilin binding activity; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2_DOMAIN_1; 1.
 DR PROSITE; PS50004; C2_DOMAIN_2; 2.
 DR PROSITE; PS50106; PDZ; 1.
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT DOMAIN 372 491 12 X 10 AA TANDEM APPROXIMATE REPEATS OF
 P-A-K-P-Q-P-Q-P-X.
 FT ZN_FING 523 547 C4-TYPE (POTENTIAL).
 FT ZN_FING 1010 1033 C4-TYPE (POTENTIAL).
 FT DOMAIN 2351 2362 POLY-PRO.
 FT DOMAIN 4442 4536 PDZ.
 FT DOMAIN 4653 4752 C2 DOMAIN 1.
 FT DOMAIN 4968 5059 C2 DOMAIN 2.
 FT VARSPPLIC 4876 4880 TKPTN -> SKRRK (in isoform 2).
 /FTID=VSP_003930.
 FT VARSPPLIC 4881 5085 Missing (in isoform 2).
 /FTID=VSP_003931.

FT	MUTAGEN	4668	4668	4668	D->A: COMPLETE LOSS OF CALCIUM-BINDING AND CALCIUM-DEPENDENT PHOSPHOLIPID BINDING ACTIVITY.
FT	MUTAGEN	4674	4674	4674	D->A: COMPLETE LOSS OF CALCIUM-BINDING AND CALCIUM-DEPENDENT PHOSPHOLIPID BINDING ACTIVITY.
FT	MUTAGEN	4688	4688	4688	V->S: SMALL INCREASE IN AFFINITY FOR CALCIUM.
FT	MUTAGEN	4688	4689	4689	VM->SS: 10-FOLD INCREASE IN AFFINITY FOR CALCIUM.
FT	MUTAGEN	4689	4689	4689	M->S: INCREASED AFFINITY FOR CALCIUM.
FT	MUTAGEN	4690	4691	4691	VV->SS: 10-FOLD INCREASE IN AFFINITY FOR CALCIUM.
FT	MUTAGEN	4692	4693	4693	QN->AA: MODERATE INCREASE IN AFFINITY FOR CALCIUM.
FT	MUTAGEN	4694	4694	4694	A->S: NO EFFECT ON CALCIUM-BINDING ACTIVITY.
FT	SEQUENCE	5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;			
FT	Query Match	3.0%; Score 451; DB 1; Length 5085;			
FT	Best Local Similarity	18.1%; Pred. No. 2e-08;			
FT	Matches	587; Conservative 445; Mismatches 1101; Indels 1116; Gaps 139;			
QY	276	ECVKPPLEVPDEWQCEVCVAKVP-----	-----GVTCVAEIQKPKYIRHEPI	322	
DB	1117	EAEPK-----VPEK-----ETASIEKTPMVT	TDQKLESEGKSKVSALPEKPEEKA	1168	
QY	323	GYDRSRKRYWFLNRLLIEDTENENEKKIYWTY	STVKVQLAELIDCDKDY-WEAE-LCKI	380	
DB	1169	SADKKERK-----PPAEKPPLEKKPI-----	-----PVDKLLPPEAKPLSSE	1207	
QY	381	LEEMREEIHRHMDITDLTNKARGSNKSF	FLAAANEIILESIRAKKGIDN-----	430	
DB	1208	GEKHILKAHVOIPE-----	-----EPTGKVAAGAGEEQPDSRPEALP	1249	
QY	431	-----VKSPEETEDK-----NETENDSKDAE	KNREEDFQSLKSDDKTDDPDEQKSE	482	
DB	1250	GATPLPLKAGEKERA	VAQPOAGSGSKDQGERSK---EKTEKE-EDKSDTSSQPKSP	1305	
QY	483	EP-TEYGDGKNSVANI	GNNTT-----NATSETSISEG-----RSPVG	520	
DB	1306	QGLSDGYSDDGISGSLGEIPSLPSDEKDLLK	GLKDKSFSQESSSPSADLAKLESTVL	1365	
QY	521	CLSETPDSSNMAEK-----KVASELPQD-----	-----VPEE-----	547	
DB	1366	SILEAQASTLVGEKA	EKKTPQKISPEKPODQKQTASSETLDTITSEEEIKESQEKVS	1425	
QY	548	PNKTCES-----	-----SNTSATTTISIQPNLENSNS-----SSELNSSQS	582	
DB	1426	PKKDSQGFPSRKEHKEKPELVDDLSP	RASYSVDSVDSSESENSESVVRRKRRTSIGSSS	1485	
QY	583	ESAKAADDPENGERESH	TPVSIQEEIVGDFTSKSTG-----ELS-----	622	
DB	1486	DEYKQEDSOGSGEEDF-----IRKQIEMSA	DEADSGSEDEEFIRSQLKEISGVGSQKR	1541	
QY	623	-ESPGAGKAGSGSTRIITRL-----	-----RNPDKSLQSKSQQA	657	
DB	1542	EEAKGKGVAGKHRLTRK	STSFDDAGRRHSHWDEDETTDESPELKFRETQSQE--	1599	
QY	658	AAAHANKLFKEGKEVLV	VNSQGEISRLSTKKVIMKGNINNYFKLGQEGKYRYVINOYS	717	
DB	1600	-----SEELVVGAGGGLRRFKT---	IELNSTIAD-----KYSSESQAK	1635	
QY	718	TNSFALNKQHREDHDKRRLAHK	FCLTTPAGEFKWNG-----SVHGSKVLTISLRITITQ	773	
DB	1636	TILY-----FDEEPELEME-SL	TSDSPDRSRGEGSSSLHASS-FTPGTSPTSVSS	1683	
QY	774	LENNIPSFPHPNWASHRANWIKAV	QMCSPREFALALAILCAVAPVVMPLPIWREFLGH	833	
DB	1684	LDSDSSPSHKKGESQ-----QKARH	SHGHPLLPTIE-----	1718	
QY	834	TRLHRMTSTERE---EKVKVKKKEQ	EBEETMQATWVYTFPVKHQVWKQGEYRYVT	890	

Db 1719 -----DSSEBELREEBELLEKEQORELEBQOQKSSSSKKKKDDDELRAQRERR--- 1769
Qy 891 GYGWSWISKTHVRFVKPLPGNTNVNVRKSLGKTGN-NMDENMDESDDKRCR----- 943
Db 1770 -----PKTP-PSNLSPIDASTFEELROAEAEELHRSSCSYSPSIE 1811
Qy 944 -SPKKIKIEPDS--EKDEKGSDAAGA-----DONEMD-----ISKIEKKDDQVKEILD- 991
Db 1812 SDPEGEISPEIKIEVQVKLPAASVLSYPTDEQSVNQKEGVQKALKSABEYEMQMOK 1871
Qy 992 -----SDSDKPKKEEPMEVDDDMKTESHVNCOESSQOVVVNVS-EGFHLRTSVKK 1040
Db 1872 PHKYKAPPAANREVEFEKPLY--GGMLIEDYI--YESLVEDYNGSVDSGLL----- 1921
Qy 1041 KTKSKLDGLERRIKOFTLEKORLEKIEKGIGIG-----KTS 1082
Db 1922 -TROEQNGFWQGRG---EQRVLOEQIYDDPMQKISDLQKEFEYESLHVVVQEDI 1976
Qy 1083 TNSKNKJESPIV-----TKAKEGCSDS-----MROESPNNANDQPEDLIOGC 1127
Db 1977 VSSSVIIPESHEIVDLGSMWMTSEKKLLDADSAYELMRQOVQVTDGSSPVQTTIGD 2036
Qy 1128 SOSDSSV--LRMSDPSTHNTKLYPKDRVLDD-----VSIRPETKCPKONSTENDIEEKS 1181
Db 2037 DMAESTLDFRVQDASLTSSIL--SGASLTDSATLSIPDVKITQOFSAELEDEYVT 2094
Qy 1182 DLASRGQBPKSKTKGNDDFFIDDSKLASADDIGTLCKNKKPLQIEESDTIVSSSKSALH 1241
Db 2095 DYTREIQD-----IAHESLIITYSEPSESA 2120
Qy 1242 SSVPKSTNDRDATPLSRAMDFEGKLGCDSESNLENSDSTVTSQDSEEDMIYVONSNES 1301
Db 2121 TSVPPSDPTSLTSSISSV-----CTTDSPPVTTLDSLTIVYTEPADVMTFKDSEE 2172
Qy 1302 ISEQPTRE-----ODVEV-LEPLKCELVSGESTGNCEDRLPVKGTENGKKPSQOKKLE 1355
Db 2173 ISSYTFPGSIIDYDEPISVLD-----RTIMPESRTN-EDRVLV---SFSGMAPSVVESVG 2224
Qy 1356 ERPVKNCSQOIKLNTDKKNENRESEKKQORTSTFOINGKDNKPKIYKLGCEKKEISE 1415
Db 2225 TKPERQADTI-----STDLPISE--KDLIKGK-----KETGDGIL--EVLDAYKD 2267
Qy 1416 SRVSGNVEPKVNNINKIPIE-----NDIKSLTVKESAIRPF-INGDVTIMEDFNERS 1467
Db 2268 KREES---EALTKIS--LPEFGLAQAPSSVTAPOIKEQHVSPHSVSGKI-----SG 2314
Qy 1468 SETKSHLLSSDAEGNVRDSLETLPSTKESD---STQTTTPSASCPESNSVQVEDMEIE 1524
Db 2315 QEKPTYRLPSGLPVSTHPS-KSRFFFRSSSLDISAQPPPPPPPPPPSPS----- 2363
Qy 1525 TSEVKKVTSSPITSEESNLNDIFENGLPINKNENYNGSKRTVITEVTTMTSTVAT 1584
Db 2364 -----TSSPPPTP-----LPPATSPKPTYPKPKKLAVAAVTSTIVTT 2403
Qy 1585 --ESKTVIVKEGDKQTVVSSTENCA-----KSTVTTTTTIVT-----KLSPTSGGS 1630
Db 2404 HYDALTMVEAAARSNGLPATKCAIAPPVPPKPSQIPTGLVTHRPEAIKPP----- 2458
Qy 1631 VDIIISVBOSTVWTTT--VTDLSLTGTGLVMTSVSKYSTKDRKVKLMKFSRPPKTRS 1688
Db 2459 ---IAPKFAVQIPVYTOQPTDTCPKPTGLSLTS-TMSLNLTSDYNNVPSTPSLSPHS 2514
Qy 1689 GTALPSYRKFTVTKSKSIFVL-----PNDLKLARKGGIREVPFYFNNAKPAIDI 1740
Db 2515 NKSSPRY-----SKSLMDTVVITLPSPEGTPTDS-----SAAQAITS 2552
Qy 1741 WPIPSRPTFTGITHWRYRLQTVKSLAGVSLMLRLWASLRWDDMAKVPDGGSGRTETISE 1800
Db 2553 WPLGSPPKDL-----VSLETVFSV-----VPPMTSTEPSASQ 2585
Qy 1801 TEITTTETIKRDRVGP--YGRIFEYCIKIIICIGVPE-TPKEPTTPORKGLRSS--ALR 1855
Db 2586 PTLTYSGALGTFSTVPANTASLFQ-----TVPSLTQFLPAEASKPEVSAVSSAVPSVA 2639
Qy 1856 PKRPETP-----KQTG--PVI-----IETVWABEELELWEIRAFAE 1899
Db 2640 PRSVSIPPPPEALDRHOYKENGKPLGLIGDAIDLRTIPKSEKVTKECMDL-----SAS 2694
Qy 1890 RVEKEKAQAVQOAKKLEQOKPPIVIATSTTSPSSSTSTTSPQAQVWVAPISGVSVTGT 1949
Db 2695 AMDYKROTTANEVYRQISAVQPSIIINLSAASSLGTPTVMTSKTVAVVVCTDITTIYTTGT 2754
Qy 1950 --KMWLTTKVGSPTAVTFQON-----KNFHQTATVWKQCSNSGVVVOQKVGLIPS 2001
Db 2755 ESQVGIHATVSPLOLITSKHTLPYKPSQAFTIRDEAPIN-----L 2799
Qy 2002 STGFSQOFTFSFQPRATVIRPNTSGSGTTSNSQVITGPQI-----RPMVTVIRPLQ 2056
Db 2800 SLGPSAQAVTLAVTKPVTPPVGVNTGTSTLSQGVADGEVDLSTSKSHRTVVTMDSES 2859
Qy 2057 QSTLIGKALI---RTPVMVQCAPOOVMTOIIRGPOPVSTAVSAPNTVSTP----- 2103
Db 2860 TSNVVTKIIIEDEKPVDLTAGR-RAVCCDMVYTLTLPFRSCTAQAPATPLPDRFGYRDDH 2918
Qy 2104 -----GOKSLTSATSTSNIOSSASQPPRQOQGVKLTMAQLTOLTQCHGNGOGLT- 2153
Db 2919 YOYDRSGPYGYRGI-----GGMKPSMSDTNLPAGHFFYKSKNAFDYSGGTGAADVLT 2972
Qy 2154 -----VVIQOGQOTTGOL---OLI-----POQVT--VLPQPGQOLMOAAM---PNG 2191
Db 2973 GRVSTGEVMDYSSKTTGYPETROVIGSVGISTPQYSTARLTTPPPQYGVGSYLRSNG 3032
Qy 2192 TVQREFLTATTATTATTTTSTTAAG--TGFQROSKLSPOMQVHQDKTLPPAQSSV 2250
Db 3033 VVYSSVATPIPTFAITQPGSIFSTVRLDLSGIPPTDAMTSLSALHQSQMPMSYRIT 3092
Qy 2251 GPAKAQQTAPQARSARPOQTOPQSPAQPEVQTOPQEVQTOPQTTVSSHVSEAPTHAQSKP 2310
Db 3093 GASET-----DIAVTGIDINASLOT-----ITMETLPAETMDSVP 3127
Qy 2311 QVAAQOSPOSNWQOSPVVRVQSPQTRIPSTPSQLSPGQOSQOVOTTTSQPIQPHSTL 2370
Db 3128 TLTATSEVSEVSGEESTLLIVPDEK-----QOQOLDL-----ERELL 3166
Qy 2371 QIPSOQOPOQOPOVQOSTQTLSGGOTLNOVSVSS-----PSRQLOIQOQPOQVIAVPQL 2425
Db 3167 ELEKIKOORFAEELEWERQETORFREQEKIMVQKLELOSMKQHLLYQOEEERQAQFMM 3226
Qy 2426 QQQVQVLSQISQVVAQIAQOOSGVPOQIKQLPIQ100-----SSAVTHOIQNVVT 2478
Db 3227 RQETLAQOOLQLEQIQLOQOQ---LHQOLEQKRLQIYQYNYDPSTSGTSSPOTTTEQAILE 3283
Qy 2479 VQAASVQEQLOQVQOQLRDOQKQKQOQIEINVNTPSKLLIKVELIQKQ----- 2526
Db 3284 QOYAATE-----GSQFWATEDATTASTVVAIEIPQSGQWTVQSDGVQ 3328
Qy 2527 -----VVMKHNVAIHLKOKKMTPAERENORMIVCNQV-----MKYIIL- 2567
Db 3329 YIAPPGLISTVSEIPLTDVVVKEEKQPKKRSAGKVRQYDEMGESVADDPNKKIVDS 3388
Qy 2568 --KIDKEE---KQAAKKRKEESVE-----OKRSKQNA----- 2595
Db 3389 GVQTDDEADKSYASRRRTTKSVDTSVQTDDEQDDEWDMPSRSRRKARTGKYGDSTAE 3448
Qy 2596 -----TKLSALL-----FKHKEQLRAELIKKRALLDKLOQ-- 2625
Db 3449 GDKTKPLSKSVSAVQTVAEISVQTEPVGTIRTPTISIRAVRDAKVEIKHISAPEKTYKG 3508
Qy 2626 -IEVQEELKRLD-----KIKKEKDMOL-----AOATAVAACP----- 2658
Db 3509 SLGCQETDSTDSQPPYLGATSPPKDKKRPPLGIGYSSSHLRADPTVOLAPSPKSPKV 3568
Qy 2659 ---PVTVPVLP-----APPAPPPS----- 2673
Db 3569 LYSPTISPLSPGNALEPAFVPVPEKPLPDDISPKQLHLPDMAKVPPASPASKMMQMSDP 3628

QY 2674 -----pppppgvqhtgllst-pt 2690
Db 3629 KPIsPTADESSRAFFOYSEGFTTKGQSTMTASQTOKKKVKKRTLPNPPPEEYSTGTQSTYST 3688
QY 2691 LPVASOKR-----KRE---EEDSSSKSKKKMISTTSKTKKDKL-- 2729
Db 3689 MGTASRRRCRTWTMARAKILQIDRLDLVERESAKLRKKQAELEDEEKEI--DAKLRY 3746
QY 2730 -----YCKTPTYDESKFYI-GCDRCQNYHGRVCVILQSEAEI----- 2767
Db 3747 LEMGINRRKEALLKEREKREKRAYLQGVAEORDY-----MSDSEVSSTRSRVESQH 3797
QY 2768 -----IDEIVCPQCTEDAMTVLTPLTEKDYGLKRLVLSLOAHKMAWPFLE 2815
Db 3798 GVERPRTAPQTEFSQIIPPTQTEAQLVPPTSPYQYQS-----S 3838
QY 2816 PVPNDAPDYGVYKIKPMDLATWEORVORR--YYEKL-----EFVADMTKIFDNCRY 2867
Db 3839 PALPTQAPTY-----TQSHFQQQTLHQVSPYQYQPTQFQAVATMFTFOAQ 3888
QY 2868 NPDSPPFYQ 2876
Db 3889 TPTPQPSYQ 3897

RESULT 4
BA2A_HUMAN
ID BA2A_HUMAN STANDARD; PRT; 1878 AA.
AC Q9UIF9; O05336; O96H26;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-I interacting protein 5) (TTF-I interacting protein
DE 5) (TIP5) (bWALp3).
GN BAZ2A OR TIP5 OR KIAA0314.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-testis;
RX MEDLINE=20130112; PubMed=10662543;
RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
RL "A novel family of bromodomain genes";
RN [2]
RP SEQUENCE OF 332-738 FROM N.A.
RC TISSUE-Lung;
RA Jansa P., Grummt I.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 639-1878 FROM N.A.
RC TISSUE-Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RL "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RN [4]
RP SEQUENCE OF 1035-1878 FROM N.A.
RC TISSUE-Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI. May serve a specific role in maintaining or
CC altering the chromatin structure of the rDNA locus (By
CC similarity).
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
CC (nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Co-localizes with the basal RNA
CC polymerase I transcription factor UBF in the nucleolus.
CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in most tissues
CC analyzed, including heart, brain, placenta, lung, skeletal muscle,
CC kidney and pancreas.
CC -!- SIMILARITY: BELONGS TO THE NAL FAMILY.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB032254; BAA89211.1; -;
DR EMBL; AF000422; AAB60864.1; -;
DR EMBL; AB002312; BAA20773.1; -;
DR EMBL; BC008965; AAH08965.1; -;
DR HSSP; Q92831; I891.
DR GENE; HGNC:962; BAZ2A.
DR MIM; 605682; -;
DR GO; GO:0005731; C:nucleolus organizer complex; NAS.
DR GO; GO:0003677; F:DNA binding activity; NAS.
DR GO; GO:0030528; F:transcription regulator activity; NAS.
DR GO; GO:0006338; P:chromatin modeling; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 4.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; SM00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 4.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
KW Nuclear protein; Repeat; DNA-binding.

FT	DOMAIN	525	577	MBD.		
FT	DNA_BIND	622	634	A.T HOOK 1.		
FT	DNA_BIND	643	655	A.T HOOK 2.		
FT	DOMAIN	821	883	DDT.		
FT	DNA_BIND	1159	1171	A.T HOOK 3.		
FT	DNA_BIND	1377	1389	A.T HOOK 4.		
FT	ZN_FING	1649	1699	PHD-TYPE.		
FT	DOMAIN	1783	1853	BROMODOMAIN.		
FT	DOMAIN	633	772	LYS-RICH.		
FT	DOMAIN	666	765	COILED COIL (POTENTIAL).		
FT	DOMAIN	1185	1250	GLU-RICH.		
FT	DOMAIN	1263	1384	PRO-RICH.		
FT	DOMAIN	1732	1735	POLY-ARG.		
FT	CONFLICT	574	574	L -> V (IN REF. 2).		
FT	CONFLICT	700	700	L -> Q (IN REF. 2 AND 3).		
FT	CONFLICT	720	720	H -> Q (IN REF. 2 AND 3).		
FT	CONFLICT	727	738	SKAEKEGKTQ -> KIKKKKKKKK (IN REF. 2).		
FT	CONFLICT	785	785	K -> R (IN REF. 3).		
FT	CONFLICT	951	951	P -> L (IN REF. 3).		
FT	CONFLICT	1005	1006	GR -> EG (IN REF. 3).		
FT	CONFLICT	1035	1037	IAA -> GTR (IN REF. 4).		
FT	CONFLICT	1163	1163	G -> S (IN REF. 1).		
FT	CONFLICT	1166	1166	R -> L (IN REF. 1).		
FT	CONFLICT	1172	1172	S -> F (IN REF. 1).		
FT	CONFLICT	1178	1178	L -> F (IN REF. 1).		
FT	CONFLICT	1202	1202	A -> V (IN REF. 1).		
FT	CONFLICT	1292	1292	P -> L (IN REF. 1).		
FT	CONFLICT	1295	1295	L -> F (IN REF. 1).		
FT	CONFLICT	1313	1313	P -> L (IN REF. 1).		
FT	CONFLICT	1407	1410	MISSING (IN REF. 4).		
FT	CONFLICT	1416	1416	R -> P (IN REF. 1).		
FT	CONFLICT	1541	1541	R -> P (IN REF. 1).		
FT	CONFLICT	1571	1571	E -> K (IN REF. 1).		
FT	CONFLICT	1616	1616	V -> I (IN REF. 1).		
FT	CONFLICT	1622	1622	E -> Q (IN REF. 1).		
FT	CONFLICT	1629	1629	Q -> H (IN REF. 1).		
FT	CONFLICT	1636	1636	Q -> H (IN REF. 1).		
FT	CONFLICT	1739	1739	R -> K (IN REF. 1).		
FT	CONFLICT	1754	1754	G -> R (IN REF. 1).		
SQ	SEQUENCE	1878 AA;	208480 MW;	400970CA68234317 CRC64;		

Query Match 3.0%; Score 447.5; DB 1; Length 1878;
Best Local Similarity 18.6%; Pred. No. 8.9e-09;
Matches 398; Conservative 297; Mismatches 763; Indels 677; Gaps 89;

QY	1082	STNSSKNLSESVITKAKGECQSDSMRQSPNANNQPEDLIQGCQSDSSVLRMSDPS	1141
DB	103	SANPGSNLKDPPLLSQFSGG-----QYPLNGILGSRPSS-----PS	140
QY	1142	HTTN-----KLY-----PKDRVLDDVSIRSPTKCPKQNSIENDIEEKVSDLASRQOE	1189
DB	141	HTNLRAGSOKFWANGTHSPMGLNFDQSLEYD---SFPDQ-----FEEVCS-----GTH	187
QY	1190	PTKSKYKGNDFDIDSKLASADDIGT-LICKNNKPLIQESDTIVS-----SKSALH	1241
DB	188	PDEAAEKEMTSVVAEN-----GTGLVCSLE---LEEOPELKCMGYNGSVPSVSLH	236
QY	1242	SSVPKSTNDRDAPLSRAMDFEGKGLGCDSESNSTLENSDTSVIOQSDSEDMIVQNSNES	1301
DB	237	QEVSVLVPD-----PTVSCID-----DPSHLPQLEDTPILS-----	268
QY	1302	ISEQFRTREODVEVLPLKCELVSGESTG-----NCDRLPVKTEANGKKPQSOOKKL	1354
DB	269	-----EDSLEPFNSLAPEVPSGGLYGIDDTFELMGAEDKLP-----L	304
QY	1355	EERPVNKCSQIOLKNTTDKNNENRESEKKGORTSTFOINGKDNKPKIYLKGECLKEIS	1414
DB	305	EDSPVISALDCPSLNNAT-----AFSLADDQ-----	332
QY	1415	ESRVVSGNVEPKVNNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSSETKSHL	1474
DB	333	-----TSTISIFASPTSPPVILGESVLQD-----NSFD-----L	359

QY	1475	LSSDAEGNYRDSLETLPTSKESDSIQTTTPSASCPESSNVQVEDMEIETSEVKVKTSS	1534
DB	360	NGSDAE---QEEME---TQSDFFPSLTPA---PDQSSTIQLHP---ATSPAVSPTTS	407
QY	1535	PITSEESNLNDFIDENGLPINKNNVNGESRKIVITVMTSTVATESKVLKVEK	1594
DB	408	PAVSLVSPAASPEIS-----PEVCPAASTVV---SPAVSFVS	443
QY	1595	GDKQTVVSSSTENCAKSTVTTTTVTKLSTPSTGGSDIISVKEQSKTVVTTVTDTSLT	1654
DB	444	PASAVLPVAVSLEVPILTASTVSPKASPVTSAPAAFTASPAKNDVSSFLETTADVEEIG	503
QY	1655	TGGLTVTSMTV-SKEYSTRDKVKL---MKFSRPKTKRSGT-----AL	1692
DB	504	EGLTASGSDVMRRRIATPEEVLPLQHGWRREVRIKKGSHRWQGETWYYPGCKRMKQF	563
QY	1693	PSYKFKVTKSKTSKIFVLPNDDLKKLARKGIGREVPYFNNAKPAALDINWYPPSPRTFGI	1752
DB	564	PEVIKLSRLNVHSV-----RRE--HFSF-----SPRMPVGD	593
QY	1753	TWRYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVP-----PGGSTRTETSET-----	1801
DB	594	FFEEF-DTPEGLO-----WVLSAEIIPSRIOAITGKRGRPRNTEKAKTEKVPKVK	643
QY	1802	-----EITTEIIKRRDVGPGYIGRFEYCIKILCIGVGPETPKETPTQKGLRSSAL	1854
DB	644	RGRGRPPKVKITELLNKTDRP-----LKKLEAOETLNEEDKAKIAKSKKMKRKQVQ	695
QY	1855	R-----PKRPETPKQGPVLIETWVAEELEWEIRAFARVEKKEKAQAVEQ	1901
DB	696	RGECLTTIOGAQRNKRQETKSLKHKAKEKSKAEKGTQOEKLEKVKREKKEKVMK	755
QY	1902	QAKKRL-----BOQKPTVIATSTTSTPTSTTST-----	1929
DB	756	KEKEEVTAKKPACKADKTLATQRRLEERQKQOOMEEMKKPTDMCLTDHQPLDFSRVP	815
QY	1930	-----ISPAQKVMVAPI---SGSVTTGTGM-----	1951
DB	816	GLTLPFGAFSDCLTIVEFLHSGVKVGLDFDPKADVPISGLVQEGLLCOGDSGEVQDILLR	875
QY	1952	VLTTKVGSPATVTFQONKNFHQTATVWVKQGSN-SGVVQVQKVLGIIPSSSTQSQ-OT	2009
DB	876	LLKAALHDPGFPSPYCSOLKILGEKVEIPLTRDNVSEILRCFLMAYGVEPALCDRLRTOP	935
QY	2010	FTSQP-RTAVTIRPNTSGSGGT-----TSNSQVITGPDIRGMVIR	2052
DB	936	FOAOPPOQAAVLAFFPVHELNGSTLIINEIDKTLSSSYKKNKWIVEGLRR-----LK	990
QY	2053	TPLOQSTLIGKALIIRTPVMVQPCAPQOVN-----TQIRGQPVSTAVSAPNTVSTPGQKS	2107
DB	991	TVLAKRT-GRS-----EVMGRPEECLGRRRRSRIMEETSGMEEEEESIAAVPGR-	1042
QY	2108	LTSATSTSNIOSSAQPPRQOGQVKLTMAQLTQLTQGHGNGQLTVVYIOGOGTQGLQ	2167
DB	1043	---GRDGEVDATASSIEPELERQIEKLSKRLFFRKLLHSSQMLRAVSLQDQRYRRYV	1099
QY	2168	LIP--QGVTVLPGPQQQLMQAMPNGT-----VQRFLETPLATTTATTTTTSPTA	2219
DB	1100	VLPLYLAGFVEGTGNLVPVEVIKKETDLSKVAHASLNPAFLSMKMLAGSNNTASSPA	1159
QY	2220	ACTGEOROSKLSPOQVHQDKTLPPAQSSSVGPAKAQPTAQPSARPOPTQPSQPAQE	2279
DB	1160	RARGPRKTK-----PGSMQPRHLKSPVRGODSEQPQAOLQPE	1197
QY	2280	VQ-----TOPEVQTOVTVSSH---VPSAOPHQAQSKPQVA-----ASQPSNVQOG	2324
DB	1198	AQLHAPAOPOLOLQLOSHKGFLEQGSPLSGOSQHDLSQSAFLWSLQTSQSHSLLS	1257
QY	2325	QSPVRVQSPQTRIPRSTPSQLSPQSQVQVTTTSQIP--IQPHTSLQIPISQOGQSQP	2382
DB	1258	SS-VLTPOSSPKLOPA-PSQ--PPEEPEDAESSPDQALWFNISAMQPCNAATPPP	1313

Qy 2375 QGQPOQ-----SQPOVQSSSTOTLSSGQFLNQVSSPSRPOLIQOQ-----POQP 2418
Dy 2975 YSOARMVSSLSPPGAGSVLRSSNGVSVVATPSTFAITTPGSGFSTVRDLPTLQ 3034
Qy 2419 VI-AVPOLO--QQVQVLSQTSQVVAQIQAOQSGVPOOIKQLPIQIQOQSSAVQTHQION 2475
Dy 3035 TIDSVPSLSLQQNQPLPSYSFLTTWAEKDASTTLDIETGLPPLFLESIAPEP-----TN 3091
Qy 2476 VVTQAAASVOEQLOQVQOQLRDQ-----QOKKKQOQIEINNVTPSKLLIKVIEIQKQVWM 2529
Dy 3092 LIPATASEVYTDV-----IEDEVALIITAPEEGKQOQLDL-----ERELLELEKIKQO--- 3138
Qy 2530 KHNAVIEHLKOKKSMTPAEREENQRMVVCNOVMKYILDKIDKEKQAAKKRKEESVEQK 2589
Dy 3139 ---RFABEELEW-----ERQEIQRF-----REQEKEMVQK 3164
Qy 2590 RSKQNAVTKLSALLFKHKEQLRAE--ILKRALLOKLOIE-----VOELKRLDKIKKEDL 2644
Dy 3165 KLEELQSMKHLLLFQOEEERQAOYMMKQETLAQOQLQLOEQFQOQLQOOLHOOLESQKTRQI 3224
Qy 2645 MQ 2646
Dy 3225 YQ 3226
RESULT 6
ID K167_HUMAN STANDARD; PRT; 3256 AA.
AC P46013:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antigen KI-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schluster C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO
CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY
CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL
CC CHROMOSOMES.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P46013-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P46013-2; Sequence=VSP_004298;
CC -1- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS
CC PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL
CC CYCLE, WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.
CC -1- SIMILARITY: Contains 1 FHA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X65550; CAA46519.1; -;
DR EMBL; X65551; CAA46520.1; -;
DR EMBL; X94762; CAA64388.1; -;
DR PIR; A48666; A48666.
DR Genew; HGNC:7107; MKI67.
DR GK; P46013; -;
DR MIM; 176741; -;
DR GO; GO:000074; P:regulation of cell cycle; TAS.
DR InterPro; IPR00253; FHA.
DR Pfam; PF00498; FHA; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS00006; FHA_DOMAIN; 1.
KW Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;
KW Alternative splicing; Polymorphism.
FT DOMAIN 27 76
FT DOMAIN 1000 2928 16 X 122 AA APPROXIMATE REPEATS.
FT REPEAT 1000 1112 1.
FT REPEAT 1122 1234 2.
FT REPEAT 1244 1356 3.
FT REPEAT 1366 1477 4.
FT REPEAT 1487 1598 5.
FT REPEAT 1608 1720 6.
FT REPEAT 1730 1842 7.
FT REPEAT 1851 1964 8.
FT REPEAT 1974 2086 9.
FT REPEAT 2096 2204 10.
FT REPEAT 2214 2326 11.
FT REPEAT 2335 2447 12.
FT REPEAT 2457 2569 13.
FT REPEAT 2579 2689 14.
FT REPEAT 2699 2808 15.
FT REPEAT 2818 2928 16.
FT NP_BIND 3034 3041 ATP (POTENTIAL).
FT VARSPPLIC 136 495 Missing (in isoform Short).
FT VARIANT 3150 3150 /FTid=VSP_004298.
FT VARIANT 3217 3217 T->S (IN dbSNP:111106).
FT VARIANT 3217 3217 /FTid=VAR_014858.
FT VARIANT 3217 3217 K->E (IN dbSNP:8473).
FT VARIANT 3217 3217 /FTid=VAR_014859.
SQ SEQUENCE 3256 AA; 358741 MW; 578F8C51BED42517 CRC64;
Query Match 2.9%; Score 439.5; DB 1; Length 3256;
Best Local Similarity 18.3%; Pred. No. 3.1e-08;
Matches 579; Conservative 432; Mismatches 1103; Indels 1049; Gaps 148;
Qy 300 VPGVTDCVAEIQKNKPYIRHEPIGYDRSRKRYWFLNRLRIIEEDTENENKKIWY----- 354
Dy 221 VP-TTQCLDNSKKNE-----SPFWKLYESVKKELDVKSQENVLQYCRKSG 265
Qy 355 ----YSTKVLQAEILDCLDKYWEAEICKILEEMREETHRHMDTDTNKAQSNKSL 410
Dy 266 LQTDYATEKESADGLQ-----GETQLL-----VSRKSRPKSGSGGHAVA 304
Qy 411 AAAN-EELLESIRAKKGDIDNVKSPET-----EKQKNET-----ENDSKDAENR 455
Dy 305 EPASPEQELDQNKGRDVSQVTPSKAVGASFLYEPAKMKTPVQVYQQQNSQPKHKNK 364
Qy 456 EEF-----EQDSLEKD-----SDDKTPDDDPQGGKSEPTGYDKNQSVS--ANLGDNTN 504
Dy 365 DLYTTGRRSVNLCKSEGFAGDKTLTPRKLTNRTPAKVEDAADSATKPKENLSSKTRG 424
Qy 505 A--TSEETSPSE---GRSPVGCCLSET-----PDSSNMAEK-----KVAEELPDQVPE 546
Dy 425 SIPTDVEVLPTETETIHNEPFLTLWTQVERKIQKDSLKPEKLTGTAGQMGSLPGLSSV 484
Qy 547 EPNKTCSSNTS-----ATTTSIQPNL--ENSSNSSELSSQSESAAK----- 588
Dy 485 DINNFGDSINESEGIPLKRRRVSGFGLHRLPELFDENLPPTPLKRGAPTAKKSLVMHTP 544


```

Db      2384 GKAMDTKPAVSDEKNINFTFVQVQLDLGNLPGSKRQPPPKKAEALDLVGPKEK 2443
Qy      2383 -QVSSQTTLSSGQTLNQVSVSPSRPQLQIQPQPQVIAVP-----OLQOOQVQLSQT-- 2435
Db      2444 FQTPGHTEESMTDDKITEVSCSPQSPESFKTSRSKQRLAIPLVKVDKKEPLAVSLKTR 2503
Qy      2436 OSQVVAIQAGQSGVPOIKL--OLPQIOOSSAVQTHQIONVYVYVQAASVQSQQLRVQQ 2493
Db      2504 TSGETTQTHTEPTGDSKSIKAFKSPKQILDPA--SVTGSRRQLRTRKKKARA 2555
Qy      2494 LRDOOQKK-----QOOIEINVT-----PSKLLIKVELIQ-----KOVVMKH 2531
Db      2556 LEDLVDFELFSAPGHTESMTDKNTKIPCKSPPELTATSTKCPTRPKVEKKE 2615
Qy      2532 NAVIEHLKOKKSMT-----PAEREENQRMVQCNVMKYILDKIDKEKQAARKKREES 2585
Db      2616 LSAVERLTQTSQSTHTHKEPASGDEGIKVL-----KQRAKKKPNPVE 2658
Qy      2586 VEQKRSKQNA-----TKLSALLFKHKEQLRA-----EILKKRA 2618
Db      2659 EEPSSRRPRAPKEKAQPLEDLAGFTELSETSGHTQESLTAGKATKIPCESPPLEVDVTTA 2718
Qy      2619 LLBKDLQIEQEEELKRLDKIKKKDLMLAQ-----ATAV 2653
Db      2719 STRHLTRVQK-----VQVKEEPSAVKFTQTSGETTDADKEPAGEDKGIKALKESAKQT 2773
Qy      2654 AAPCPVPVPLPAPPAPPPPP-----PPPGVQHTGLLSTPTLPFVASQKRKREBEK 2705
Db      2774 PAPAASVTGSRPRAPRESAQAIEDLAGFKDPAAGHTESMTDDTKTKIPCKSSPELED 2833
Qy      2706 SSKSKKKMISTSTETKDKTKLYCTICKTPYDESKFYICDRQCNWYHRCVGIQSEA 2765
Db      2834 TATSKRRPRTRAKQVEKKEE--LLAVGK-----LTQTSQ 2866
Qy      2766 ELIDEXVPCQSTEDAMTVLTPTLTKEDYGLKRLVLSLQAHKMWPFLEPVDPN-DAPD 2824
Db      2867 E-----THTDKEPVGEGKGTKA-----FKQAKRNDAED 2897
Qy      2825 YGVVKEPM-----DLATMEERVOR-RYYEKLTFEVAD 2856
Db      2898 VIGSRQPRAPKEKAQPLEDLASQFELSQTTPGHTTEELANGAAD 2940

RESULT 7
ANK2_HUMAN
ID   ANK2_HUMAN          STANDARD;          PRT;   3924 AA.
AC   Q01484; Q01485;
DT   01-APR-1993 (Rel. 25, Created)
DI   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN   ANK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   TISSUE=Brain stem;
RX   MEDLINE=91302466; PubMed=18330053;
RA   Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT   "Isolation and characterization of cDNAs encoding human brain
RL   ankyrins reveal a family of alternatively spliced genes.";
RL   J. Cell Biol. 114:241-253(1991).
RN   [2]
RP   REVISIONS.
RA   Carpenter S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain stem;
RX   MEDLINE=94075409; PubMed=8253844;

```

```

RA      Chan W., Kordeli E., Bennett V.;
RT      "440-KD ankyrin: structure of the major developmentally regulated
RL      domain and selective localization in unmyelinated axons.";
RL      J. Cell Biol. 123:1463-1473(1993).
RN      [4]
RP      SEQUENCE OF 463-495 FROM N.A.
RX      MEDLINE=92009921; PubMed=1833308;
RA      Tee W.T., Menninger J.C., Yang-Peng T.L., Francke U., Sahr K.E.,
RA      Lux S.E., Ward D.C., Forget B.G.;
RT      "Isolation and chromosomal localization of a novel nonerythroid
RL      ankyrin gene.";
RL      Genomics 10:858-866(1991).
CC      -1- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC      elements. Also bind to cytoskeletal proteins.
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=3;
CC      Name=1;
CC      IsoId=Q01484-1; Sequence=Displayed;
CC      Name=2;
CC      IsoId=Q01484-2; Sequence=VSP_000267, VSP_000268;
CC      Name=3;
CC      IsoId=Q01484-3; Sequence=VSP_000268;
CC      -1- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC      CELLS THROUGHOUT THE BRAIN.
CC      -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC      AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC      AND FUNCTION (POTENTIAL).
CC      -1- SIMILARITY: Contains 23 ANK repeats.
CC      -1- SIMILARITY: Contains 1 death domain.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X56957; CAA40278.1; -
DR      EMBL; X56958; CAA40279.2; -
DR      EMBL; Z26634; CAB42644.1; -
DR      EMBL; M37123; AAA62828.1; -
DR      PIR; S37431; S37431.
DR      HSPR; P42771; IDC2.
DR      Genew; HGNC:493; ANK2.
DR      MIM; 106410; -
DR      InterPro; IPR002110; ANK.
DR      InterPro; IPR000488; Death.
DR      InterPro; IPR000906; Z05.
DR      Pfam; PF00023; ank; 23.
DR      Pfam; PF00531; death; 1.
DR      Pfam; PF00791; ZU5; 1.
DR      PRINTS; PR01415; ANKYRIN.
DR      SMART; SM00248; ANK; 22.
DR      SMART; SM00005; DEATH; 1.
DR      SMART; SM00218; ZU5; 1.
DR      PROSITE; PS50088; ANK_REPEAT; 20.
DR      PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR      PROSITE; PS50017; DEATH_DOMAIN; 1.
KW      Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW      Phosphorylation.
FT      REPEAT 63 92 ANK 1.
FT      REPEAT 96 125 ANK 2.
FT      REPEAT 129 158 ANK 3.
FT      REPEAT 162 191 ANK 4.
FT      REPEAT 193 220 ANK 5.
FT      REPEAT 232 261 ANK 6.
FT      REPEAT 265 294 ANK 7.
FT      REPEAT 327 360 ANK 8.
FT      REPEAT 331 360 ANK 9.
FT      REPEAT 364 393 ANK 10.
FT      REPEAT 397 426 ANK 11.
FT      REPEAT 430 459 ANK 12.

```

```
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.
FT REPEAT 694 723 ANK 20.
FT REPEAT 727 756 ANK 21.
FT REPEAT 760 789 ANK 22.
FT REPEAT 793 822 ANK 23.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A.
FT REPEAT 1857 1867 REPEAT A.
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A.
FT REPEAT 1892 1902 REPEAT A.
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH.
FT VARSPLIC 1039 1039 Q -> QFLGKLHLPTAPPLNAGESLSVRIQLGPPGTYK
(in isoform 2).
FT VARSPLIC 1444 3528 /FTID-VSP_000267.
Missing (in isoform 2 and isoform 3).
FT CONFLICT 475 476 GO -> VSP_000268.
FT CONFLICT 971 971 I -> PE (IN REF. 4).
FT CONFLICT 3581 3582 QY -> S (IN REF. 1).
FT CONFLICT 3586 3586 QY -> HA (IN REF. 1).
FT SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;
2.9%; Score 439.5; DB 1; Length 3924;
Best Local Similarity 17.7%; Pred. No. 3.8e-08;
Matches 539; Conservative 421; Mismatches 1115; Indels 967; Gaps 125;
QY 196 ESDKEYHVLPOEADPYGVENKIKVLQVDFLTNNAREELMSEGVIQYDDHCR 255
||||| : : : : : ||| : : : : : ||| : : : : :
D 1427 ESDQOEEDIDMTSEKNDTESTSV-LKSHLVNE--VPVLASPDLLSEVSEMKQDLIK 1483
||| : : : : : ||| : : : : : ||| : : : : :
QY 256 -----VCHKLGLLCCETCSAVYHLECVKPPLEVPEDMQCEVCVAHKVPGVTDCV 307
||| : : : : : ||| : : : : : ||| : : : : :
D 1484 MTAILTVDSDKAGSIKVELYKAA-----EEFGEFPE-----IVERKEDLEKY 1529
||| : : : : : ||| : : : : : ||| : : : : :
QY 308 AEIQNKPKYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKKIWIYSTRVQLAELIDC 367
||| : : : : : ||| : : : : : ||| : : : : :
D 1530 NEILSGTGCTDES-SVQSSRSERGLIVEEWIVVSDEETEARQ----KAPLEITE-YPC 1583
||| : : : : : ||| : : : : : ||| : : : : :
QY 368 LKDDTWEALCKILMEERIEHRMD-----ITEDLTNKGARSKNSFLAAANEETLESIR 422
||| : : : : : ||| : : : : : ||| : : : : :
D 1584 VE-----VRIDKEIKGVBEKSDTGLVNYLTDLNT-----CVPLPKQLQTVQ 1626
||| : : : : : ||| : : : : : ||| : : : : :
QY 423 AKKG-----DI--DNVKSPETEKD-----KNETENDSKDAEKNREE-- 457
||| : : : : : ||| : : : : : ||| : : : : :
D 1627 DRAGKKCEALAVRSESEKGDIPDETQSTQKQKPSLGKIKKPYRKLKEKQKQKEBGL 1686
||| : : : : : ||| : : : : : ||| : : : : :
QY 458 ---FEDQSLEKSDDKTDDDDPEQGKSEPTVEVGDKGNSVSANLGDNNTTNATSETPSE 514
||| : : : : : ||| : : : : : ||| : : : : :
D 1687 QASAEKAEULKGSSESLGEDP--GLAEPPL-----PTVKATSPLI 1725
||| : : : : : ||| : : : : : ||| : : : : :
QY 515 GRSPVCLSETPDSNMA---EKKVASELP-----QDVPEPNKTCSSNTSATTTSTQ 565
||| : : : : : ||| : : : : : ||| : : : : :
D 1726 EETPIGSKDKVKALQKRVYEDQKGRSKLPIRVKGEDVPK---KTHRPHPAAS----- 1777
||| : : : : : ||| : : : : : ||| : : : : :
QY 566 PNLNS-----NSSSELNSQESAKAADDP---ENGERSHTPVSIQEEIVGDTFSE 615
||| : : : : : ||| : : : : : ||| : : : : :
```

```
Db 1778 PSLKSERHAPGSPSPKTERHSTLSSSAKTERHPVPVSPSSKTEKHSPVSPSAKTERHSPAS 1837
QY 616 KSTGELSPGAGKAGSGSTRIITRLNRPDSKLSQKSOQVAAAHAANKLFKEGKEVLV 675
||| : : : : : ||| : : : : : ||| : : : : :
Db 1838 SSSKTEKHSP-----VSPSTK--TERHSPVSTKTERHPVPVSPSG-----KTDKRPVP 1883
||| : : : : : ||| : : : : : ||| : : : : :
QY 676 VNSQGEISRSLTKKEVIMKGINNINNYKLGQGGKRVYHNOYSTNSFALKKHQHRDHDK- 734
||| : : : : : ||| : : : : : ||| : : : : :
Db 1884 ----SPSGRTEKHPPVSPGRTEKRLPVPSPGR-----TDKHQPVSTAGTK 1924
||| : : : : : ||| : : : : : ||| : : : : :
QY 735 RRHLAHECLTPAGEFKWNGSVHGSVKVLTITLRLTITQLENNIPSSFFHFNWASHRAW 794
||| : : : : : ||| : : : : : ||| : : : : :
Db 1925 EKHLIP-----VSPSGKTEKQPPVSP-----TSSTERIETMSVREL----- 1960
||| : : : : : ||| : : : : : ||| : : : : :
QY 795 IKAYQCMCKPREFALALAILCAVAKVPVVMPLPIWREFLGHTRLHRMTSTEREKEKVK-KK 853
||| : : : : : ||| : : : : : ||| : : : : :
Db 1961 MKATOSGODPSKHKTGLFEHKSQAKQ-----QPQEKGVKRVK 1998
||| : : : : : ||| : : : : : ||| : : : : :
QY 854 EK-----KQBEETMQOATWVKYTFPVKHQVWQKGEYRVTYGGWSWISKTHYRVFP 908
||| : : : : : ||| : : : : : ||| : : : : :
Db 1999 EKGPILTQREAKTENQTI-----KRGQRLPVTG----- 2027
||| : : : : : ||| : : : : : ||| : : : : :
QY 909 KLPGNTNVNYRKSLEGTKNNDENDESDDRKCSRPKKIKIEPDEKDEVKGSAAKGA 968
||| : : : : : ||| : : : : : ||| : : : : :
Db 2028 --TAEKRGVRVSSIGVKK--EDAAGGKEKVLSHKIPFVQSVPEESHRESEVPKEMA 2083
||| : : : : : ||| : : : : : ||| : : : : :
QY 969 D-QNEMDISKITEKK-DQDVKELDSDSKPCKEPMEDVDDM-KTESHVNCQESSQVDV 1025
||| : : : : : ||| : : : : : ||| : : : : :
Db 2084 DEQDMDLQISPRDKTSTDFSEVIKQLEEDNDKYQOFRLESEETKAQLHLD-----QVLT 2138
||| : : : : : ||| : : : : : ||| : : : : :
QY 1026 VNVSEGFHLRTSYKK-----KTKSSKLDGLLE-----RRIKQFTLEE 1062
||| : : : : : ||| : : : : : ||| : : : : :
Db 2139 SPFNTTFPL--DYMKDEFPLALSQSGALDGSSLSKNEGAVAGSPCGSLMGTPQISSEE 2196
||| : : : : : ||| : : : : : ||| : : : : :
QY 1063 KORLEKIKLEGIGIKGISTNSSKNLSSEPVITKAKEGQCSQSMRQSPQSNANNDQ--- 1119
||| : : : : : ||| : : : : : ||| : : : : :
Db 2197 SYKHE-----GLAETPETSPLSFSPKKEEQGTEKESKTKTTTTEIRSEKHEP 2247
||| : : : : : ||| : : : : : ||| : : : : :
QY 1120 -PEDLIQCSQSDSVLRMSDPSTNKLYPKDRVLDVDSRSPKPKONSLEND--- 1175
||| : : : : : ||| : : : : : ||| : : : : :
Db 2248 TTKDITGSEERGATV---TEDSETSFQKEATLG-----SPKDTSPKR---QDDCTG 2296
||| : : : : : ||| : : : : : ||| : : : : :
QY 1176 -----TEEKVSLASR--GOEPTKSKTKGNDFFIDSKLASADDITGLICK 1219
||| : : : : : ||| : : : : : ||| : : : : :
Db 2297 SCVALAKETPTGLTEEAACDEGQRTFGSSAHKTOT-----DSEAQES-----TATSD 2344
||| : : : : : ||| : : : : : ||| : : : : :
QY 1220 NKKPLIQEESDTIVSSSKSALSHSSVPKSTNDRDATPLSRAMDFFEGKLCDCSESNTLNS 1279
||| : : : : : ||| : : : : : ||| : : : : :
Db 2345 ETKALPLPEASVKTDGTGTSKPQGVIRSPQGLELALPSK-----DSEVLASAVDD 2394
||| : : : : : ||| : : : : : ||| : : : : :
QY 1280 SDTVSIQSDSEEDMIVQNSNESISQFTRDODVEVLEPLKCELVSGETGNCEDRL--- 1336
||| : : : : : ||| : : : : : ||| : : : : :
Db 2395 SLAVSHKDSLEASPVLEDNSS-----HKTPDLSLEP-----SPLKESPCRDLSLESS 2439
||| : : : : : ||| : : : : : ||| : : : : :
QY 1337 PVKGTGEANGKPS-----QKKLEERPVNKCSD 1364
||| : : : : : ||| : : : : : ||| : : : : :
Db 2440 PVEPKMAGIPIPSHPPLPAVAKTELLTEVASVRSRLLRDPDGSADDSLEOTSLMESSG 2499
||| : : : : : ||| : : : : : ||| : : : : :
QY 1365 QIKLKNNTDKNNENRESEKKQRTSTFQINGKDNKPKIYKLGECCLKIEISRVVSGNVE 1424
||| : : : : : ||| : : : : : ||| : : : : :
Db 2500 KSPLSPDTPSSEVSEVTPKTTDVST-----PKPAVTHECAEEDD-----SENGE 2545
||| : : : : : ||| : : : : : ||| : : : : :
QY 1425 PKVNNINKLIPENDIKSLTVKESATRPFGINGDVIMEDFNERNSSSEKSHLLSSSDAEGNY 1484
||| : : : : : ||| : : : : : ||| : : : : :
Db 2546 KK-----RFTPEEEMFKVTKIMFDELEQEAQKRDYKKKEPKQESS-----SSSDPDADC 2597
||| : : : : : ||| : : : : : ||| : : : : :
QY 1485 RDSLETLSTKESDSTQTTTPSASCPESNSVNOVDMETSEVKKVTs----- 1533
||| : : : : : ||| : : : : : ||| : : : : :
Db 2598 SVDVDE-PKHVGGGDESGVPVLVTSERKVSSESSEPELAQLKKGADSGLLPEPVRV 2656
||| : : : : : ||| : : : : : ||| : : : : :
QY 1534 ---SPITSEESNLNSDFI-----DENGLPINKNNVNGES-----KRRT 1570
||| : : : : : ||| : : : : : ||| : : : : :
Db 2657 QPPSPLPSSMDSNSPEEVQFQPVVYSKYVTFKMNEDTQEPGKSEEDKSESHLAEDRHA 2716
||| : : : : : ||| : : : : : ||| : : : : :
```


RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEvan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wuzley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield J.S.N., Krzyzinski M.I., Skalska U., Smalus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5].
RP SEQUENCE OF 4405-5147 FROM N.A.
RA Kalicki J., Elliott G.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9Y6V0-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC VSP_003926, VSP_003927;
CC Note=No experimental confirmation available;
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DRH domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y19188; CAB60727.1; -;
CC EMBL; AC004903; AAD20936.1; -;
CC EMBL; AC004886; AAD21789.1; -;
CC EMBL; AB011131; BAA25485.1; -;
CC EMBL; BC001304; AAR01304.1; -;
CC EMBL; AC004082; AAB97937.1; -;
CC PIR; T00634; T00634.
CC HSSP; P04410; 1A25.
CC Genew; HGNC:13406; PCLO.
CC MIM; 604918; -;
CC GO; GO:0005856; C:cytoskeleton; NAS.
CC GO; GO:0045202; C:synaptic junction; ISS.
CC GO; GO:0005509; F:calcium ion binding activity; ISS.
CC DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
CC DR GO; GO:0005522; F:profilin-binding activity; ISS.
CC DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001565; Synaptotagmin.
CC PRINTS; PR00360; C2DOMAIN.
CC PRINTS; PR00399; SYNAPTOTAGMN.
CC SMART; SM00239; C2; 2.
CC PROSITE; PS00499; C2_DOMAIN1; 1.

DR PROSITE; PS00004; C2_DOMAIN2; 2.
KW Calcium/phospholipid-binding; zinc; Metal-binding; Zinc-finger;
KW Repeat; Alternative splicing.
FT NON_TER 1
FT DOMAIN 400 465 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-P-Q-P-X.
FT ZN_FING 499 523 C4-TYPE (POTENTIAL).
FT ZN_FING 969 992 C4-TYPE (POTENTIAL).
FT NON_CONS 1010 1011 POLY-PRO.
FT DOMAIN 2300 2325 PDZ.
FT DOMAIN 4391 4442 C2 DOMAIN 1.
FT DOMAIN 4344 4633 C2 DOMAIN 2.
FT DOMAIN 5031 5121 C2 DOMAIN 2.
FT VARSPLIC 4404 4404 S -> SGNGLGIRIVGGKEIPFGHSGEIGAYIAKILPGGSAE
FT QTKLMGEG (in isoform 2).
FT /FTid=VSP_003923.
FT K -> KPTDGGKVVVSHPTGETQ (in isoform 2).
FT /FTid=VSP_003924.
FT G -> GOVMVVONAS (in isoform 2).
FT /FTid=VSP_003925.
FT TAHSK -> SKRRK (in isoform 2).
FT /FTid=VSP_003926.
FT Missing (in isoform 2).
FT /FTid=VSP_003927.
FT SQ SEQUENCE 5147 AA; 563537 MW; CDS084990498CD3C CRC64;
Query Match 2.9%; Score 438.5; DB 1; Length 5147;
Best Local Similarity 18.2%; Pred. No. 5.6e-08;
Matches 554; Conservative 379; Mismatches 938; Indels 1171; Gaps 135;
Qy 275 LECVKPL--BEVP-----EDWQCEVCVAHKVPGVTCVAVIQKNKKYIRHEPIGY 324
Db 1076 LEKVETLSMEKIPPMVTDDQESKLE-----KDKASALQEKKLPE----- 1119
Qy 325 DRSRRKYWFLNRLIIEEDTENENKKIWIYSTVKVQLAELIDCLDKDYWEALCKILEP 384
Db 1120 -----EKKLPEEEKIRSEKK-----PLLEE- 1141
Qy 385 REEHRHMDITEDLTNKGKSNKSFLLAAANEELIESIRAKKGDIDNVKSPETEKKNE- 443
Db 1142 -----KKPTPEDKLLPEAKTSAPEE---QKHL--LKSQVQIAEKLK 1181
Qy 444 -----TENDSKDAKNREPEF---QSLKESDDKT-----PDDPEQK 480
Db 1182 RVAPKTVOEGKQPTKMEGLSPGTPQSLPKD-DKTKTKIKEQPPCTAKPDQEKEDK 1240
Qy 481 SE-----EPTEVGDKGNS---VSANLGNTT-----NATSEETSPSE 514
Db 1241 SDTSSSQPKSPQGLSDTGYSSDGISSILGIPSLIPTDEKDIKGLKDSFSQESSPSS 1300
Qy 515 G-----RSPVGCILSETPDSSNMAEKVASELPQDV-PEEPNKTCESSNTSATTSTQPN 567
Db 1301 PSDLAKLESTVLSILEAQASTLADSEKSEKTKTPHEVSEPEQ-----KDQEKTSQSL 1353
Qy 568 LENSNSSELNSSESASAKAADDPENGERSHTPVSTOEIVGDTSEK-----STGELSE 623
Db 1354 LEITISEEIKESQERKDTFKDQSQDIPSKDHKEKSEFVDDITRRREPYDVSVEESE 1413
Qy 624 SPGAGKASGSTRIITLRNP---DSKLSQLKSOQVAAAHANKLFKEGKVLVNVNSQ 680
Db 1414 SE-----NSVPQKRRRTSVGSSSDDEYKQDSQSGSGEEDFIRKQ-----IIMSAD 1461
Qy 681 EISRLSTKKEIVKMGNNINNYFKLQGEQKRYVYHNYQYSTNSPALNKHQHRDHRRLAH 740
Db 1462 EDASGSEDEDFI-----RNQLKEIS--SSTESQKKE----- 1490
Qy 741 KFLTLPAGEFKWNGSVGSKVLITSTLRLLTTQLENNIPSPFFPNWASHRANIKAVQM 800
Db 1491 -----ETKGGKITAGK----- 1502
Qy 801 CSKPREFALALALECAVKPVVMLPIWREFLGHTRLHRMTSIEREEKEKVKKEKKEE 860
Db 1503 -----HRLTRKSTSTIDE-DAGRRHSHWDEDD 1529


```

Db 3265 SQGWTVSGDGVTOYI-----APP-----GILSTVSEIPLTDVVVYKEEKQPKKR 3308
QY 2705 DSSSKSKKKMISTTSKETKDKTKLYCICKTPYDESKFYIGCDRCQNNYHGRVCVILQSE 2764
Db 3309 SSGAKVRGQ-----YDDMGENTDDP-----RSF 3332
QY 2765 ABLIDEYVCPQCQSTEDAMTLVPLTERKDYGLKR-----VLSRLQAHKMAWPFLEVPDP 2819
Db 3333 KKVDSGV-----QTDD-----EDATDSYVSRRRTKKSVDTSVQ-----TDD 3371
QY 2820 NDAPDYGVYKPEMDLAWEERQVR--RYEKLTFEVDMTK 2859
Db 3372 EDQDEW-----DMPTRSRKARVKGYSMTS--ADTKK 3403

RESULT 9
BAZA_MOUSE
ID BAZA_MOUSE STANDARD; PRT; 1850 AA.
AC Q91YE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-1 interacting protein 5) (TTF-1 interacting protein
DE 5) (Tip5).
DE DE
GN BAZA2A OR Tip5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423468; PubMed=11532953;
RA Strohm R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,
RA Laengst G., Grummt I.;
RT "NORC-a novel member of mammalian ISWI-containing chromatin remodeling
RT machines.";
RL EMBO J. 20:4892-4900(2001).
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI. May serve a specific role in maintaining or
CC altering the chromatin structure of the rDNA locus.
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
CC (nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Co-localizes with the basal RNA
CC polymerase I transcription factor UBF in the nucleolus.
CC -!- SIMILARITY: BELONGS TO THE WAL FAMILY.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ309544; CAC69992.1;
CC MGD; MGI:2151152; Baza2a.
CC DR GO: 0005731; C:nucleolus organizer complex; ISS.
CC DR GO: 0003677; F:DNA binding activity; ISS.
CC DR GO: 00030528; F:transcription regulator activity; ISS.
CC DR GO: 0006338; P:chromatin modeling; ISS.
CC DR GO: 0006355; P:regulation of transcription, DNA dependent; ISS.
CC InterPro: IPR006637; AT_hook.
CC InterPro: IPR001487; Bromodomain.
CC InterPro: IPR004022; DDT_dom.
CC InterPro: IPR001739; Methyl-CpG_bind.
CC InterPro: IPR001965; Znf_PHD.

```

```

DR Pfam; PF02178; AT_hook; 4.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 4.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
KW Nuclear protein; Repeat; DNA-binding.
FT DOMAIN 505 557 MBD.
FT DNA_BIND 602 614 A.T HOOK 1.
FT DNA_BIND 623 635 A.T HOOK 2.
FT DOMAIN 800 862 DDT.
FT DNA_BIND 1137 1149 A.T HOOK 3.
FT DNA_BIND 1351 1363 A.T HOOK 4.
FT ZN_FING 1623 1673 PHD-TYPE.
FT DOMAIN 1755 1825 BROMODOMAIN.
FT DOMAIN 613 738 LYS-RICH.
FT DOMAIN 647 774 COILED COIL (POTENTIAL).
FT DOMAIN 1006 1012 POLY-GLU.
SQ SEQUENCE 1850 AA; 205597 MW; E5DD4FED1D46DAE1 CRC64;

Query Match 2.8%; Score 422.5; DB 1; Length 1850;
Best Local Similarity 18.1%; Pred. No. 6.6e-08;
Matches 390; Conservative 295; Mismatches 772; Indels 693; Gaps 82;

QY 1082 STNSKNLSSESPV-----TKAKGCGQSDSMRQSQSNA 1115
Db 61 SANPNNNKADPPLLSQLQPGQYPLNGILGNRQSPSSHNTRLRAGSEFWANGTQSPMG 120
QY 1116 NNDQPEDLIQGCQSDSSVLNRSDPSHTTNKLYPKDRVLDDVSRSPETKCPKQNSIEND 1175
Db 121 LNFDSQELYDSDPDQNFVMPNGPPSFFTS---POTSPMLGSSI---QTTFAPSQD-VSSD 173
QY 1176 I-----EEKVSDLASRGOEPTKSKTGNDFFIDSKIASADDITGLCKNKKPLIOBE 1228
Db 174 IHPDRAARKELTTSVVAENG-----TGLVGSLEEEQELKMC 211
QY 1229 SDTIIVSSSKSALHSSVPKSTNDRDATPLSRAMDFEGLGCDSESNSTLENSSDTVSIQDS 1288
Db 212 GYNGSVSVSESLHQEVSVLVPD---PTVSCLD-----DPSHLPDQ 248
QY 1289 SEEDMIVQNSNESISEQFTTRQDVVEVLEPLKCELVSGESTG-----NCDRLPVKGT 1341
Db 249 LEDTPTLS-----EDSLEPFDLSAAEPVSGSLYGIDDAELMGAEDKLPLEG- 294
QY 1342 EANGKKPSQOKLEERPYNKCSDOIKLKNTTDDKKNENRESEKKGORTSTFQINGKNKP 1401
Db 295 -----NPVISALD----- 302
QY 1402 KIYLGECLEKISESRVSGNVEPKVNNIN--KIIPENDIKSLTVKESAIRPFINGDVIM 1459
Db 303 -----C-----PALSNANAFSLADDSDTSASIFVSPSPVLGSEVL 340
QY 1460 ED--FERNSETKSHLLSSDAEGNYRSLTFLTKRESSTQ-----TTTFSASCP 1510
Db 341 QDNSEGLNCSDSQEIEETQSS--NFQRLP-TEFAPDQPPSTLHPVSPFASPAASLT 397
QY 1511 ESNVNOVEDMEIETSEVKKVTSSPTSEESNLSNDFIDENGLPINKNENVNGESKRKT 1570
Db 398 ASAEISPA-----VSPVASSVP----- 415
QY 1571 VITEVTTMTSTVATESKTVIKVEKDGKQTVVSTENCNAKSTVTTTTTTTKLSTSTGGS 1630
Db 416 --PEVFVAVSPASSPALPAISLE-----ASMTTPVTSPQSGSPSPAAAA 457

```



```
Db 1884 -----EQNG-----FMPQGRE---QKIRLREQIYDDPMQKITD----- 1914
QY 1379 NRESEKKGORTSTFOINGKNPKIYLKGECLKEISRSVVGNNVEPKVNNIKIIPEND 1438
Db 1915 -----LQKEFE----- 1933
QY 1439 IKSLTVKESAIRPFGDVIEMDFNERNSETSKSHLLSSDAEGNYRDSL--ETLPSTKE 1496
Db 1934 I-----VSSVLIIPESHVIDLGSMTSTSEKK--LL--DADAAYEELMKROOMQVTDG 1984
QY 1497 SDSQTTTPSASCPSNSVQVEDMEIETS-----EYKVTSSPITSEESNLSNDFIDE 1551
Db 1985 SSLIOTMGDDMAESTLDFDRYQDASLTSSLGASLTDSTSSATSLIPDVKITQHSTE 2044
QY 1552 NGLPINKNENGVESKKKTVITEVTMTSVAVESKTVIKVEKDKQTVVSSSTENCAKST 1611
Db 2045 E-----FEDEVYTDYTRIEQIIAHESLITYSEFSATSVPPSDTSLSSSVCTTD 2100
QY 1612 VTTTTVTVKLSTPSTGGSDIISVKEQSKTVVT-----TVTDSL 1652
Db 2101 SSSPVTLDSLTVYT--EPADVITKFKDSEISTVPPGSDIVDPEDIGVSLDRTIPES 2159
QY 1653 TTTGGTLVTSM-----TVSKEYSTRDKVKLMKFSRKRKTRSGALPSYRKFVTKSTKS- 1706
Db 2160 RTNADQIMISFPGIAPSITESATKPE-----RQADTISTDLPISEKELIKGKETG 2212
QY 1707 -IFVLPNDLKLARKGGIREVYFNNAKPAIDWIPSPRPTFGITWYRLQTVKSLA 1765
Db 2213 DGIILEVLDAYKDKRESEAEALTKISL---PETGLAPTSPSSQT-----KEQGPSHSVS 2263
QY 1766 GVSMLRLWLASLRWDDMAAKVPPGG---STRTETSETIITTEIKRDRDVPYGRFE 1822
Db 2264 G-----EISGQKPYRSPGGLPVSTHPSKSHPPFRSSSLDISAQPPP----- 2307
QY 1823 YCIRKIICPGVETPKETPTORKGLRSSALRPKRPETPKQ-----TGPVLIETWA 1875
Db 2308 -----PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 2356
QY 1876 EELELEWEIRAFARVEKEKAQAVEQO-----AKRLEQQKPTVIATSTTSPT----- 1923
Db 2357 -----AHADAIPTEVATAARRSGLPATKICAAAAPPVPPKPSIPTGLVETHRP 2406
QY 1924 SSTTSTISPAQKVMVAPISGVITKMWLTWKVGSPTVTFQONKFNHQTFTATWKQG 1983
Db 2407 EASKPPIAPKPAVEIPVITQKTDT---CPKPTGLPLTNSMLNL----- 2449
QY 1984 SNSGVVQVQKVLGIIPSTSTGTSTQOTFTSPQR-----TATVTIRPNTSGSGGTTNS 2036
Db 2450 -----VTSADYK-----LPSPTSLPSHNSKSPRYSKSLMETYVVIPLSEPPTDSSAA 2501
QY 2037 QVITGPIRPGMTVIITPLQOQSTLGLKALIRTPVMVQPGAPQVMTQI-IRGQPVSTAVSA 2095
Db 2502 QAITST-----WPLGSPPKDLVSLVTFVS---VPPMTSTEIPSAQPTLTSTGA 2547
QY 2096 PNTVSTPG-QKSLTSATSTSNLOSASQPPRQOQOVKLTMAQLTQLTQGHGNGGLTV 2154
Db 2548 LGTFSVTPAVTASLFTVPTSLTQFLPAEASKSEVSAVSAVSP-----RSVSI 2599
QY 2155 VI-----OQOGQTTGLOLIPQGVTVLPGLPGQOL----- 2183
Db 2600 PIPPEPLADLRHOKYKNGKPLIGDAIDLRTIPKSEKVKTEKMDLSASAMDVKRTTAN 2659
QY 2184 -----MOAMPNGTVQREFLTPLATTATTASTTTTSTTAAAGTEQROSKLSPOM 2234
Db 2660 EYVRRQISAVQPSIINLSAASSLGTPVMTDSKIVAVTCT-DTIYTTGTE-----SQV 2712
QY 2235 QVHQDKTLPPAQSSVGPKAQPTAQPSARPOQTPQOSPAQ-----PEVQ----- 2281
Db 2713 GIEHAVTSPLOLTTST---RHTELQYKRPSSQAPPFIRMIRDEAPINLSLGPSTQAVTLAVTKP 2769
QY 2282 -TOPEV-----OTQTTVSSHV-PSEAQPTHQSSKPVAAQSQPSNV-----QQQSPV 2328
Db 2770 VTVPVGVTVNGWTDSTISOGITDGEVVVDLSTSKSHRTVTVMTDESTSNVVTKIIEDEKPV 2829
```

RESULT 11

```
TRX_DROVI
ID TRX_DROVI STANDARD; PRT; 3828 AA.
AC Q24742; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```


Db 1632 --ELLSEQPPWFONETKACTDALEDWFBSCGVEELKESPTTYAEHHTASQAPRTGLLOI 1689
QY 1153 VLDVVSIRSPETKCPKONSIENDIEBKVSDLASRGOEPTKSKTK-----GNDFIDDSKL 1207
Db 1690 PLDDVDDLG---GC-----AVKRLDTRVCLFCRKSGEGLSGEARLLYCHDCWVH--- 1738
QY 1208 ASADDIGTLCKNKKPLIOEESTIYSSSKSALHSSVPKS-----TNDRDAT----- 1254
Db 1739 -----INCAMSAEVEFEID-----GSLQNVHSAVARGRMIKCTVCNRRGATVGCNVK 1786
QY 1255 -----PLSRAMD-----FEGKLGCDSBSNSTLE-NSRDTVSIOSSSEEDMIVONSNES 1301
Db 1787 SCGEHYHPCARTIDCAFLTDKSNYCPAHARNALKANGSPSYIESNFV-----SRPV 1840
QY 1302 ISOQFTRRODVEVLEPLKCEL-----VSGESTGNCEDR-----LPVKGTBA-----NG 1345
Db 1841 YVELERKRK---KLIIPAKVQFHIGSAVRQLGSIVPRFSDSPEAIVPINFLCSRLYWSS 1897
QY 1346 KRPQOQKLEERP-----NKCSDQIKL---KNTDKKNENRESEKK----- 1386
Db 1898 KEP---WKIVEYTVRTTIONSYSSTLTLDAGRNFYDHTNPNC SLVOLGLAQIARHSSL 1954
QY 1387 QRTSTEQINGKONKPIYKGECLKEISESRVSVGNVEPKVNNIKIIPENDIKSLTVKE 1446
Db 1955 ARSDLLDTWAE-FPNYSYPAD---ENTEE-----EFQ-QNADLLPPE--IKD-AIFE 1999
QY 1447 SAIRPFING-----DVIM-EDFNERN-----SSETKSHLLSSDAENGRDSTLETLPTSTKE 1496
Db 2000 DLPHELLDGLMIDFEMWEDLGKTELFAMSEQSGKGTWATSQAG--ASVILCDETRN 2057
QY 1497 SDS-TOTTPSASCPSNSVNOV-----EDMEIETSEV-KVTTSSP-----I 1536
Db 2058 SNSLAKHLVLSNCTASNPVDDAMLCAARSSSOEKECGDVLKKTDTAPTNRWPKLDGGSV 2117
QY 1537 TSSEENLSNDFIDENGLPINKNENNGESKRTVITEVTTMTSTVATESKTVIKVEKGD 1596
Db 2118 AAFKRKLSKNIAE--GVLLSLNQ-----RSKK-----EMATVAGITR-- 2153
QY 1597 KOPVSVSTENCASVTVTITTTTKLSTPSTGGSVDIISVKEQSK-----TVVTTT-VT 1649
Db 2154 ROSVCGSSSELPAGSATMRTKSP-----TWSAAKCLFEKNESREAPKLTIMQMDGVD 2206
QY 1650 DSLTT-----TGGTLTSMTVSK-----EYSTRDKVK-----LMKSRPKRTS 1688
Db 2207 DSITEYRIIGSDGNLSTAFTGOVKERCOCYTRNYDFSQRHLGSCPEMSTSESEAT 2266
QY 1689 GTALPSYRKFVTKSKKSIFVLPNDLKK-----LARKGGIREVPYFNYNKAPALDIW 1741
Db 2267 GTAQLSAESL-----NELOKALAAATLSNTGGL-----NY-----LQTSF 2302
QY 1742 PYPSPRPT---FGITWRYRLQTVKSLAGVSL-----MLRLWLASL 1778
Db 2303 POVONLALQCGVQVLOGLOTLQ-LQPSGLNGFFLSQPNAAQATSNGNDVLQLYANS 2361
QY 1779 RWDMAAKVPPGGSGSTRSETSEITEITTELIRRDVGPYIRFEYICIRKILCPIGPVETP 1838
Db 2362 Q---NLAAANL--GGGFLTQPTMSTQAPQLI-ALSTNPDG-----TQ 2398
QY 1839 KETPTPQRKGLRSSALR---PKRPETPKOTGPVIIETWAAEELELWEIRAFARVEKEK 1895
Db 2399 OFIQLPQNGATYOLLQTAAPLRCNATYOT---LQATNSDKKIVLLFLEA-GDPLQEVV 2453
QY 1896 AQAVEQ-----QAKRLEQOQKPTVIATSTPTSSTSTTSTISPAQKV 1936
Db 2454 TOAAQQAATAAAHOKOLKSGHGVKPIQAKLOGOQOQOQHOOHQOQOQOQOQOQOQ 2513
QY 1937 MYAPISGVTGTGWLLTKVSGPATVTFQOQKNFHOTFATVWKQCSNSGVVOQKVL 1996
Db 2514 QOTPIVAQGGTGTOLLGONLLQ-LQLFQSNQAQ-PQOQQLLPLPQOQN----- 2561
QY 1997 GIIPSTGTQSQ-----TFTSFQPTATVTIRPNTSGSGGT----- 2033
Db 2562 -IISFVVGSGSONOPLYISITPNDFKPQOQTSTPTFLTAPGGGATFLQDASGNMLT 2620
QY 2034 -----SNSQVITGPQIRPGMTVIRTPLOQSTL-----GKAIIRTPVMVQP----- 2073
Db 2621 TAPANSGLQMLTG-QLOTOPOVIGTLIQOTLQTTGADGTQTATATAQOPLILGATGGGT 2679
QY 2074 -----GAPOQVMTOIIRGPV-----STAVSAPNTVSS-T-PG---QKSLTSATST 2114
Db 2680 TGLEFATAPVILAT---QPMYIYGLTIVONTVMSSQSFVSTAMPQVLSQNSFSFATTT 2735
QY 2115 SNIQSSASOP---PRP-----QOGOVKLTMAQLTQLTQGHGGNGLT 2153
Db 2736 QVFQASKIEPIVDLPAGYVVLNNAVDSAGNTSWLQSQSTQATDDATAQLLQ-NAGFOFQT 2794
QY 2154 VTIQGGOTGTQQLIPQGV-----VLPQPGO-OLMNA 2187
Db 2795 TPTTSTQOTMSTDYAPPLVVTAKVPPVAQMKRNTNANKSPISVLSKVOPQOQSVVNVK 2854
QY 2188 MPNGTV-----QRFLPTLATATTASTTTTT-----VSTTAAG--T 2222
Db 2855 LPTNVIOOQO 2914
QY 2223 GEORQKSLSPOMOVHODKT--LPPAQSSSVGP--AKAOPOTAPSPARPOTOPQSPA-- 2276
Db 2915 GSTCGAPPSTASRPLQKKNLIRPIHKVEVKPKIMQAPKLATSAASHQHHQOQSPAAI 2974
QY 2277 -----QPEVOTQPEVOTQTTVSSHVPSEAPTHAQSSRPQVAAASQOPQ 2319
Db 2975 NOVAKVALLQORLAPAPQO-QOEPQOEEQHLHQOQOQOQOQOQOQOQOQOQOQOQOQO 3033
QY 2320 SNVQGSVP-----RVQSPQTRIRPSTPQSPQOQOQOQOQOQOQOQOQOQOQOQOQO 2363
Db 3034 L-LRAQOPIISIVNTAPQAAOTFVIRPALQAQAPITQLOEQOQOQOQOQOQOQOQOQO 3092
QY 2364 --IOPHTSLQIPSQ-GOPOSQOPOVOSTOTLSSSGOTLNQVSVSSPSR-----POLQIQO 2414
Db 3093 ARLQRYASNSLPTNVNPLQOQRCASANNSSNNTQONSTIINSRPTNRVLPQOQOQO 3152
QY 2415 POP-----QVIAVP-----OLOQOQOVLSQIQSOV---AQIQAOQSGVPOQIKLQLP- 2462
Db 3153 PTPLSNDVVVQSPPTPKPIEPPVAGASTOKPIVKCYAQLQEQSGPYETELKTNITLDNL 3212
QY 2463 QQSSAVQTHQIQ-----NVVTQAAASVQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2505
Db 3213 EOTNITTMOLOQO 3272
QY 2506 IEINVTAPSKLLIKVEILQOVVYMKHNAVIE-----HLKOKKS-----MT 2545
Db 3273 MEMVVDNGFOLTSN---ESCLLEKHGFNVEAVPMDTEDHYASKNKGSGGAAEGIGQVD 3328
QY 2546 PABREEN-----QRMIVCN-----QVMKYILDKIDKEE---KQAAKRRKR 2582
Db 3329 DAEDEDDDDDFSLKMATSDNDHMSDSEEPVAVKEKISKILNDLNTDSDSSTATATTV 3388
QY 2583 EESVEQKRSKON--ATKLSALLFKHKEOLRA-----EILKKRALLDKLOLIE 2627
Db 3389 EASAGYQOQVVEDVLTATAGSVSTDDTTTATAEVAASYNEMAAHELOLQLOQAG 3448
QY 2628 VOBELRDLKIKKEKDLMLAOATAVAACPPTVLPAPPAPPSPPPPPVQHTGLLS 2687
Db 3449 VELDLK-----KPKLDVPOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3492
QY 2688 TPTLPVQAKRKRKEBKSSKSKKMKMISTTTSKETKDKPKLYCICTKPYDES----- 2740
Db 3493 GPHLLYEIQSDGFTYKSSSIAIEWEKVFQVARRAHG-----LTPLPEGLADMSG 3546
QY 2741 -----KPYI---CCDRCONWY-----HRCVGV-----ILOSE 2764
Db 3547 VOMIGLTKNALYIEQLPQVEKVCYKTPYKHKRNGNVSTAAGGHARTAGSNPAALAG 3606
QY 2765 AE-LID-----EYVCPQOCQSTEDAMTTLTLETKDEYGLKRLVLSLQAHKMAWP 2812
Db 3607 AESLIDYGSQOEBELQENAYECARCE-----PYVSRSEYDMFSLASRRHKQPIQV 3656


```
QY 48 ----SYCT-----ESSFRSHSTYSSTPCRKP-----RVHRPRSPILLEKDIPPLE 89
Db 1501 RPYOSWTAPITVPCPAKSGFTLSSSSSNTPSPASPLKSIWSVSTP-SPIKSTLGASTTS 1559
QY 90 FPKSSEDLMPVNEHIMVIAIYELRNFVTLRSPRFEFCAALYSQEOCTLMAEMHV 149
Db 1560 SVKISDVASP-----IRSLRTMSSPIKTVVOSPYNIQVSSGTARAPAVTATPLKG 1613
QY 150 VLLKAVIREEDTSNTTF-----PADLKDSVNSTLYFIDGMTWPEVLRVCES 197
Db 1614 LASNSTSSRTSPVTTAGSLERSSITWTPASPKNINM----- 1653
QY 198 DKEYHHVLPYQ-----EADYPYGVVENKIKVLQFLOFLTNTARELMSEGVIOYDDH 253
Db 1654 ---YSSLPEFKSIITSAAPLISSPLKSVVSKVRVDVISAKITMASSLSPPVKMPGH 1710
QY 254 CRVCHKLGDL-----LCCECTSAVYHLECVKPLEEVPEDQCEVCAHKVPCVT 304
Db 1711 AEVALVNGISPLXYASSTLINGCKATATLQ-----EKISSATNSVSVVSAAT 1760
QY 305 DCVAEIQKN-----KPYIRHEPIGYDRSRK-----YWFNLRRLLIIEEDTENENEK 350
Db 1761 DTVEKFVSTTTAMPFSLRSVVSAAFAQSLRTPSASALYTSLSGSSISATTSVTSII 1820
QY 351 KIWTYST-----KVQLAELIDCLDKYWEAELKILEEMREEIH--RHMDITEDLTNKAR 403
Db 1821 TVPYVSVVNLPEPALKLPDSNFTKSAALLSPIKTLTETHPQHFSRTSPV---- 1876
QY 404 GSNKSFLA-----AANEELIESIRAKGDDINVK-----PEETE 438
Db 1877 -KSLFLAPSALKLSTPSSLSSSOEILLKDVAMKEDLMRTAIILOTVPEKPPQPELPK 1935
QY 439 KDKNETENDSKDAEKNEEFEDQS--LEKD--SDKTDDPEQKSEEPTEVGDKGNV 494
Db 1936 EGRIDDEPFKIVEKVEDLVKVEILLKDKVCVNK--GSPKSPKS-----DKGHSP 1985
QY 495 SANLGDNTNATSEETSPSEGRSPVGLSETPSSNMAEKVASELQDVPEEPN-KTCE 553
Db 1986 E-----DOWIEFSEEI--REAR-----QQAASQSP-SLPERVQVRAKA 2022
QY 554 SSNTSATTISIQPLENSSELSSNSESASAKAADPPENGERSHTP-VSIOQE----- 607
Db 2023 ASEKDYNLTVIDYLTNDIGSSLTNLKYKPEDAKKGEGQKRVLPALQAEHLKMP 2082
QY 608 --IVGDTSEKSTGELSESPGAGKAGSGSTRIITRLRNP-----SKLS-----QLK 652
Db 2083 PASMRTSTSEKELCKMADS-----FRGTDTI--LESPDDFSQHDQDKSPLSDSGFETR 2133
QY 653 SOQVAAAHEA-----NKLPE-----GKEVLV-----NSGEISRLTKKEVI 692
Db 2134 SEKTPSAQSAETGPKPLFHEVPIPPVITETRVVHVIRSYDPSAGDVPQTQPEEVS 2193
QY 693 MKGNNVFKLGQEGKYRVVHNOYSTNSFALNKHQHREDHDKRHLAHKFCCLTPAGEFKW 752
Db 2194 PKPS-PIFMEL--EPKPTTSIKEKVFQMKASSEEDHN--RVLS----- 2235
QY 753 NGSVHGSKVLTISTLRUTITQLENNIPSSFPHPNWASHRA-----NWKAVQMCKSPR 805
Db 2236 ----KGMRVKEETHITTTTRVYHSPPG-----EGASERIBETMSVHDIMKAFQSGRDP 2287
QY 806 EFALALAILCAKVPVNLPLWREFLGHTRLRM---TSTEREKVKKKKEKQEBEET 862
Db 2288 KELAGLFHKSAVSPV-----HKSAAETSQAHAEDKNQMPKLERIIE-- 2331
QY 863 MQQATWVKYTPPVKHQVWKQGEYRVTYGGWMSWISKTHVYRVFVKLPGLNTNRYKSL 922
Db 2332 -----VHIEKNQAEPTV-----IIRTKKHPEKEMVYQKDL 2365
QY 923 -EGTKNNMDENDESRRKGRSPKKIKIEPDSKDEVKSGDAAGADQONEMDISKITEK 981
Db 2366 SRGIDNLKDFLPEKHDAFPCSE-----QOQOEELTAEESLPSYLESSRVNTPVSQE 2419
QY 982 KQO--DVKELLDSQDKPKC---EEMVEVDD-----MKTESHVNCQESSQVDVVNVSEGF 1032
```

```
Db 2420 EDSRPSSAQLISDSDSYKTKLLISQHSIEYHDELSELRGESYRFAEKMLLSEKLDVS--- 2476
QY 1033 HLRTSYK---KKTSSKLDGLLER-RIKQFTLEEKORLEKTKLEGGIKGKTGTSTNSK 1087
Db 2477 HSDTEESYTDIAGPPSELQGSDDKRSREKRIATAPKKEILSKLYKDVSVNGVKVS----- 2531
QY 1088 NLSESPVITKAKEGCQSDSMRQEQSPNANNDQPEDLIOGCSQSDSVLRMSDPSHT---- 1143
Db 2532 -----KDEHFDKVTVLHYSGN-----VSSPKHAMMR 2558
QY 1144 -----TNKLYPKDRVLDVSTRSPETTKCPQNSIENDIEEKVSDIASRGQBP-TKS 1193
Db 2559 FTEDRLDRGREKLIYEDRV--DRTVKEAEELKTEVSQFFRDKTEKLDN---ELQSPERKA 2613
QY 1194 KTKGNDFPIDDSKLASADD---IGTLCKNKK-----PLIOESDFTIVSS 1236
Db 2614 RPKNGKEYSSOSPSSSSPEKVELLTELASDENWYKAROHGPDGQGFPAEKAEPSLPSP 2673
QY 1237 KSALHSSVPKSTNDRDAT-----PLSRAMD-----FEGKLGCDS 1270
Db 2674 EKWVLS---QOTEDSKSTVEAKGSISQSKAPDGPQSGFOLKQSKLSSIRLAFEOGTHAKS 2730
QY 1271 ESNSTLENSDTS-----IQDS-----SEEDMIYQNSNESISEQ-----FTR 1309
Db 2731 KMSQEDRKSDGSRIPVKKIOESKLPVYQVFAREKQOKAIDLDPDESYSQKDFMVLTK 2790
QY 1310 BODVEVLB-----PLKCELYSGESTGNCEDRL 1336
Db 2791 DEHAQSEINIVYNDSGDNVKKQRTSEMSSKAMPDSFSEQAQDLACHITSDIATRGPDWK 2850
QY 1337 PVKQTEANG--KPSQQKKLEERPYNKCSQDIKLTNTDKKNE--NRESEKKGORTSTF 1392
Db 2851 VFTWESSGATNNKSQEKLSHLVH-----DVRENHIGHPESKVDQKNFEM 2898
QY 1393 QINGKDNKPKIYLGECLKEISESRVSGNVEPKVNINKIIPENDIISLTVKESAIRPF 1452
Db 2899 SVTERERK---LLANGSLSEIKEMTVKSPS-----KKVLYREYVVREG 2938
QY 1453 INGDVIMEDNERNSETKSH-----LLSSDAEGNVRDSELTLPSTKESDSQT 1502
Db 2939 DHPGLLDQPRRESSAVSHIPVRVADERMLSSNIPDGPCEQS--APPKHELQKLSQ 2996
QY 1503 TTPSASCPESNVQVDEMEIETSEVKVT-----SSPITSEE----- 1540
Db 2997 SSMSKETVETQHFNSIEDEKVTYSEISKVHQSVGLCPPLEETETSPTKSPDSLEFSP 3056
QY 1541 -ESNLNDPFDENGLPINKNENV-----NGESKRKTVITEVTTMTSTVATESKTVIKE 1593
Db 3057 GKESPSSDVFDHS--PIDGLEKLAPLAQTEGGKEIKTLPVVY---SFOVGKQYKEIQ 3110
QY 1594 KGDKQTVVSSSTENCASKTVTITVTKLSTPSTGGSVDIISVKBQSKVTVTVTTSIT 1653
Db 3111 QGGVKKLIISCKVQVQETRGFTYTRQOKQPPSPQSGPEDDTLEQVS-----FLD 3160
QY 1654 TTGGLTVTSMT-----VSKEYSTRDKVKLMKF-----SRPKKTRSGTALPS 1694
Db 3161 SSGKSPLTPETPSSEEVSEYFTSKTTPDSLAIYIPGKPSPIPEVSESEEBEQAKSTL-- 3218
QY 1695 YRKFTVSTKKSIFV--LPNDLKLKARKGIREVPVFNYNAPKALD----- 1739
Db 3219 ---KQTTVEETAVEREMPNDVSKDSNQRKNRNVAYIEFPPLPLDADQIESDKHHY 3273
QY 1740 -----IWPVSP-----RPTFGITW 1754
Db 3274 LPEKEDMIEVNLQDEHDKYOLAEVPVIRVQP-PSVPPGADVSSDDESIVQPVVKY 3332
QY 1755 RYRLQTV-----KSLAGVSLMLRLWLAS-----LRWDDMAAKVPPGGGSTRTE 1797
Db 3333 TFKLKEVDDEQEKPKASAEKASNOKELESNGKDNFEGGLDPSQNEIAQNGNDQSI 3392
QY 1798 TSETITTTTEIKRRD-----VGPYGIRFE-----YCIKRII 1829
```


Db 244 TSSTSESTSSSTAPATPTTTSCTKEKPTPTTTSCTKEKPTP-----PHDDTTPCT 296
Qy 1367 KLANTDKKNENRESEKKQORTSTFGINGKDNKPKIYLKGECLKEISRVVSGNVPEPK 1426
Db 297 KKKTTTSKCTKKTTPVPPTSSSTTE-----SSSAPVPTPSSSTTESSSAPVTSSTTE-- 350
Qy 1427 VNNIKIIPNDJKSLTVKESATRPINGVDIMEDFERNSETKSHLLSSDAEGNYRD 1486
Db 351 --SSSAPV--TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP 394
Qy 1487 SLETPTKESDS---TOTTPSAPCESNVQVEDMETETSEVKKVTSPTTSE--- 1539
Db 395 VPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV--TSSTTESSAPVTSSTTES 453
Qy 1540 -----EESN--LSNDFIDENGLPINKNENNGESKRKTIVTEVTMTST--VA 1583
Db 454 SSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVP 513
Qy 1584 TESKTVKVEKDKQTVVSTENCACKSTVTTTTVTVKLSTPGSGVDIISKEQSKTV 1643
Db 514 TPSSSTTESSAPAPTTPSSSTTESSAPVTSSTTESSAPVPTPSS-----STTESSTP 568
Qy 1644 VTTVTDSLTGTGLTVMVSKKE--YSTRDKVKLMKFRPKKTRSGTALPSYKRFVTK 1701
Db 569 VTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPAPTTPSSSTTESSAPVTS 628
Qy 1702 STKKSIFVLPNDLKKLARKGGIREVPYFYNNAKPALDIWPPSPRPTFGITWRYLQTV 1761
Db 629 STTESSA-----ETPKETPTQKGLRSSAL----- 1854
Qy 1762 KSLAGVSLMLRLWASLRWDDMAAKVPPGGSGTRTETSETIIT-----TEIKR 1811
Db 648 SSAPVPTP-----SSSTTESSAPVPTPSSST--TESSAPVTSSTTESSAPVTSSTTE 701
Qy 1812 RDVGPGYRPEYCIKILICIGVP-----ETPKETPTQKGLRSSAL----- 1854
Db 702 SSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESS 761
Qy 1855 -----RPKRPTPKOTGVIITVWABEELELWIRAFARVEKEKAQVQQAKKLEQQ 1910
Db 762 APVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSNTSS 821
Qy 1911 KP--TVIANTTS-----PTSTSTSTISPAQVMVAPISGVTGTGKMLTKVGS-- 1959
Db 822 APSSTPSSSTESSVPVPTPSSSTTESS-----APVSSSTTESSAPVPTPSSSN 874
Qy 1960 -----PATVTFQONKFNHOFATWVKQGSNGVQVQKVLG-----IIPS----- 2001
Db 875 ITSAPSSIFPSTTESFST-GTIVTPSSSKYPGSQTESVSTTETTVPTKTTTSVTT 933
Qy 2002 -----STGTSQ--QTFTSFQPTATVTRPNTSGSGGTTNSQVITGPQIRPGM 2048
Db 934 PSTTTTITTVCSGTNSAGETSGCSPKVTTVV-PTTTTTSVTSSTTIT-----T 985
Qy 2049 TVIRTPLQ---QSTLG---KALIRT--PVWQPCAPQVMTQIRGPVSTAVS-----AP 2096
Db 986 TVCSGTNSAGETSGCSPKTIITTVPCSTSPSETASESTTPTTPTVTVSTTVTTE 1045
Qy 2097 NTVSSTPGOKSLFASSTNSIQSS---ASQPPRQOQOV-KLTMAOLTOLTOGHGGNOGLT 2153
Db 1046 YSTSTKPGGE-ITTFVTKNIPTLYLTITAPTSTVTVNFTTTTITTVCSGTNSA-- 1102
Qy 2154 VTIQOGQTTGQQLIPQGVTVL-----PGPGOOLMQAAMPNGTVORFLFTPLATTAS 2209
Db 1103 -----GETTSGCS--PKVTVTVPCTSGTGEVTEA-----TTLVTTA--- 1138
Qy 2210 TTTTTSVTAAGTGQRQSKLSQMVHQDKTLPAPQOSSVSGPAKAQPTQAQPSARPQ 2269
Db 1139 -VTTTVTVTSTGTSAGKTTTG---YTTKSPVTVTVTLAPS----- 1178
Qy 2270 TQPSAPQVQVQPEVQVQVQVTVSSHVPSBAQPTHAQSSKPVQAAQSQSNVQGSQV 2329
Db 1179 ----APVTATNAPVTTTTECSAATNAGETTSVCSAKTIIVSSASAGENTAPSAT--- 1231

Qy 2330 VQSPSQTRIPRST--PSQLSPGOQSQVOTT---TSQPIQPHFTSLQIPSGQSQSQPOV 2384
Db 1232 --TPVTTAI-PTTVITTESSVGTNSAGETTTGTTKSIPTTYITTL-IPGNGAKNYETV 1287
Qy 2385 QSSTOTLSSQTLNOVSVSPSRPOLQIQPOPOVIAVQLOQOVVLSQIQSOVQAQIQ 2444
Db 1288 ATAT-----NPISIKTTS-----QLATTASASSVAPV 1315
Qy 2445 AOSGVPQIKLOLPQIQIOQSSAVOTHQIONV-VTVQAAS 2483
Db 1316 TSPS-----LNGPLQASGSASAVATISVPSISSTYOGAA 1348

RESULT 15

MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Cuihane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -1- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC -1- SIMILARITY: Contains 2 WVC domains.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2003, 16:10:05 ; Search time 56.2184 Seconds
(without alignments)
8207.607 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

Sequence: 1 MVSEEEEDGDAETQDSE.....KLGFASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	14971	100.0	2907	21	AA197452 Human transcriptio
2	14170	94.6	2781	21	AA157453 Human transcriptio
3	13735	91.7	2759	24	AA016418 Human nucleic acid
4	2695.5	18.0	557	22	AAB94078 Human protein sequ
5	1488.5	9.3	976	22	AB68735 Drosophila melanog
6	1385	9.3	328	21	AAB54364 Human pancreatic c
7	1030	6.9	191	24	ABU70846 Human adipocyte Se
8	982.5	6.6	238	20	AA104323 Human secreted pro
9	981	6.6	237	22	AAU19446 Human diagnostic a

10	789	5.3	149	23	AAU78845	SCAM/KRAB protein
11	781	5.2	246	22	AAU31944	Novel human secret
12	661.5	4.4	997	22	ABG62887	Drosophila melanog
13	621.5	4.2	2768	22	ABG68397	Drosophila melanog
14	613.5	4.1	5533	22	ABG5772	Drosophila melanog
15	613.5	4.1	5560	22	ABB71160	Drosophila melanog
16	567	3.8	6815	22	ABG6811	Drosophila melanog
17	562.5	3.8	3257	22	ABG67502	Drosophila melanog
18	561	3.7	3080	22	ABG64877	Drosophila melanog
19	551.5	3.7	2897	22	ABG58514	Drosophila melanog
20	503.5	3.4	4498	22	ABG58595	Drosophila melanog
21	491	3.3	2137	23	ABF39618	Drosophila melanog
22	474	3.2	3726	22	ABG63947	Staphylococcus epi
23	462	3.1	2586	22	ABG66878	Drosophila melanog
24	460	3.1	10431	24	ABU54861	Human CA125 amino
25	458.5	3.1	3266	21	AA842491	Human ORFX ORF2255
26	456.5	3.0	3664	24	ABR47592	Breast cancer asso
27	456	3.0	3021	24	ABP75873	Human secretory po
28	451.5	3.0	2283	24	ABP56876	Staphylococcus epi
29	449.5	3.0	3111	22	ABG60327	Drosophila melanog
30	448	3.0	3236	24	ABP96155	Human Ki-67 chromo
31	444	3.0	2977	22	ABG69480	Drosophila melanog
32	443	3.0	2261	24	ABU18914	Pathogen specific
33	443	3.0	2344	22	ABU37120	Staphylococcus aur
34	442	3.0	1870	24	ABJ19019	Pathogen specific
35	441.5	2.9	3696	23	ABP40235	Staphylococcus epi
36	439.5	2.9	1878	22	AAU40239	Human polypeptide
37	439.5	2.9	3256	21	AAU50976	Human cell cycle p
38	439.5	2.9	3256	24	ABU07489	Protein different
39	438	2.9	3256	23	ABG77188	Prostate adenocarc
40	435	2.9	1727	22	ABG95554	Human protein sequ
41	433	2.9	1795	22	ABG69806	Drosophila melanog
42	432	2.9	2665	22	ABG48336	Human liver peptid
43	432	2.9	2665	22	ABP28314	Human peptide #965
44	432	2.9	2665	22	ABP33490	Peptide #996 encod
45	432	2.9	2665	22	ABU18950	Protein #949 encod

ALIGNMENTS

RESULT 1
AA157452
ID AA157452 standard; Protein; 2907 AA.

XX AC AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

XX XX AA157452;

Human transcriptional regulatory factor SEQ ID NO:1.

Human transcriptional regulatory factor; TCOA1; BLAST detection;
bromo-domain; cell proliferation; cancer.

Homo sapiens.

WO957143-A1.

11-NOV-1999.

30-APR-1999;

30-APR-1998;

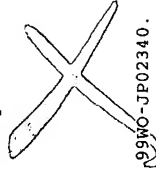
(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

Jones MH;

WPI: 2000-052940/04.

N-PSDB; AAZ39032.

Transcriptional regulatory factor containing a bromo domain and gene
TCOA1 encoding it



XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
XX
XX
SQ

Claim 1; Page 48-68; 154pp; Japanese.

The present sequence represents a human transcriptional regulatory factor containing a bromo domain. The factor interacts with proteins involved in the chromatin-mediated transcription regulatory mechanism. It binds to hSNF2H, hSNF2L and NCoA-62/Skip. It can be used for screening compounds binding to it and acting as agonists or antagonists, which are potentially useful for the treatment and prevention of cancer and other cell proliferation disorders.

Query Match 100.0%; Score 14971; DB 21; Length 2907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MYSEEEEDGDAETQSEDDDEDEDEDDSDYPEEMEDDDDDASYCTESSFRSHST 60
|||||
1 MYSEEEEDGDAETQSEDDDEDEDEDDSDYPEEMEDDDDDASYCTESSFRSHST 60
|||||
61 YSSTGRRKPRVRRSPRIEKKDIPPLEFPKSSDLMPNHEIMNVIAIYEVLRNFGTV 120
|||||
61 YSSTGRRKPRVRRSPRIEKKDIPPLEFPKSSDLMPNHEIMNVIAIYEVLRNFGTV 120
|||||
121 LRLSPFRFDFCAALVSOQCFLMAEMHVLLKAVLREEDTSNTTGGPADLKDSVNSTLY 180
|||||
121 LRLSPFRFDFCAALVSOQCFLMAEMHVLLKAVLREEDTSNTTGGPADLKDSVNSTLY 180
|||||
181 FIDGTMWPEVLAVYCESDKYHHVLPYQAEADYPGPVENKTKVLQFLVDQFLTTNIARE 240
|||||
181 FIDGTMWPEVLAVYCESDKYHHVLPYQAEADYPGPVENKTKVLQFLVDQFLTTNIARE 240
|||||
241 ELMSEGVIOYDHCRCVCHKLGLDCCETCSAYVHLECVKPPLEEVPEDEWCQECVCAHV 300
|||||
241 ELMSEGVIOYDHCRCVCHKLGLDCCETCSAYVHLECVKPPLEEVPEDEWCQECVCAHV 300
|||||
301 PGVTCVABIQNKPYIRHEPIGYDRSRKKYWFNLNRLLIEDTENENKKIWIYSTKVQ 360
|||||
301 PGVTCVABIQNKPYIRHEPIGYDRSRKKYWFNLNRLLIEDTENENKKIWIYSTKVQ 360
|||||
361 LAELIDCLDKYWEAFCLKILPEMREEIHRHMDITEDLTNKGNSKSFLLAANAEIILES 420
|||||
361 LAELIDCLDKYWEAFCLKILPEMREEIHRHMDITEDLTNKGNSKSFLLAANAEIILES 420
|||||
421 IRAKKGIDNVKSPETEKKNETENDSKDAEKNEEFEDQSLKSDDDKTPDDDPEQK 480
|||||
421 IRAKKGIDNVKSPETEKKNETENDSKDAEKNEEFEDQSLKSDDDKTPDDDPEQK 480
|||||
481 SEPTVEGDKGNSVANLGDNTNATSEETSPSEGRSPVGCLETTPDSSNMAEKKVASEL 540
|||||
481 SEPTVEGDKGNSVANLGDNTNATSEETSPSEGRSPVGCLETTPDSSNMAEKKVASEL 540
|||||
541 PDVPEPNKTCESNTSATTTISQPNLENSNSSSELNSSQESAKAADDPNGERESHT 600
|||||
541 PDVPEPNKTCESNTSATTTISQPNLENSNSSSELNSSQESAKAADDPNGERESHT 600
|||||
601 PYSIQEIVGDFTEKSTGELSESPGAGKAGSGSTRIITRLRNPDSKLSQLKSQQVAAAA 660
|||||
601 PYSIQEIVGDFTEKSTGELSESPGAGKAGSGSTRIITRLRNPDSKLSQLKSQQVAAAA 660
|||||
661 HEANKLFGKGEVLLVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNQYSTNS 720
|||||
661 HEANKLFGKGEVLLVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNQYSTNS 720
|||||
721 FALNKHQHRHEDHDKRRLAHKFLCTPAGFEFKWNGSVHGSVKLTISTLRLTITOLENNIPS 780
|||||
721 FALNKHQHRHEDHDKRRLAHKFLCTPAGFEFKWNGSVHGSVKLTISTLRLTITOLENNIPS 780
|||||
781 SFFHPNASHRANWIKAVQCMCKPREFALALAILCAVAKPVVYMLPIWREFLGHTRLHRMT 840
|||||
781 SFFHPNASHRANWIKAVQCMCKPREFALALAILCAVAKPVVYMLPIWREFLGHTRLHRMT 840
|||||

QY 841 SIEREEKEKVKKKKKQEBEETMOQATWVKYTFPVKHQVWKQKGBEYRVTCYGGWSWLSK 900
DB SIEREEKEKVKKKKKQEBEETMOQATWVKYTFPVKHQVWKQKGBEYRVTCYGGWSWLSK 900
QY 901 THVYRFVPLKPGNTNVNYRKSLIEGTKNNMDENMDSDRKCSRPKKIKIEPDESEKDVK 960
DB THVYRFVPLKPGNTNVNYRKSLIEGTKNNMDENMDSDRKCSRPKKIKIEPDESEKDVK 960
QY 961 GSDAAKAGADQNMEDISKITERKKDQDVKELLSDSDKPKCEPMEVDDDDMKTESHVNCQES 1020
DB GSDAAKAGADQNMEDISKITERKKDQDVKELLSDSDKPKCEPMEVDDDDMKTESHVNCQES 1020
QY 1021 SQVDVVVYSEGFHLRTSYKKTKSKLDGLLERRIKQFTLEEKQORLEKIKLEGKIGLKG 1080
DB SQVDVVVYSEGFHLRTSYKKTKSKLDGLLERRIKQFTLEEKQORLEKIKLEGKIGLKG 1080
QY 1081 TSTNSKNLSESPVITTKAKEGQSDSMRQEQSPNANDQPEDLIQCCSQSDSSVLRMSDP 1140
DB TSTNSKNLSESPVITTKAKEGQSDSMRQEQSPNANDQPEDLIQCCSQSDSSVLRMSDP 1140
QY 1141 SHTTNKLYPKDRLVDVDSIRSPETKCPKONSIENDIEEKVSDLASRGQEPKSKTKGNDF 1200
DB SHTTNKLYPKDRLVDVDSIRSPETKCPKONSIENDIEEKVSDLASRGQEPKSKTKGNDF 1200
QY 1201 FIDDSKLASADDIGTLICKNKKPLIOEESDTIVSSSKSALHSSVPKSTNDRDATPLSRAM 1260
DB FIDDSKLASADDIGTLICKNKKPLIOEESDTIVSSSKSALHSSVPKSTNDRDATPLSRAM 1260
QY 1261 DPEGLGCDSESNSTLENSDTSVIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK 1320
DB DPEGLGCDSESNSTLENSDTSVIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK 1320
QY 1321 CELVSGESTGNCEDRLPVKGTAEANGKPSQOKKLEERPVPVNCSDQIKLNTTDDKNNENR 1380
DB CELVSGESTGNCEDRLPVKGTAEANGKPSQOKKLEERPVPVNCSDQIKLNTTDDKNNENR 1380
QY 1381 ESEKKGORTSFQINGKDNKPKIYLGECLEKBEISRVVSGNVPEKVNINNKIIPENDIK 1440
DB ESEKKGORTSFQINGKDNKPKIYLGECLEKBEISRVVSGNVPEKVNINNKIIPENDIK 1440
QY 1441 SLTVKESAIRPFIINGDIVMEDFERNSSSETKSHLLSSDAEGNYRDSLETLPSTKESDST 1500
DB SLTVKESAIRPFIINGDIVMEDFERNSSSETKSHLLSSDAEGNYRDSLETLPSTKESDST 1500
QY 1501 QTTTFSACPSNSVNOVEDMEIETSEVKKVTSSPITSEESNLNDFIDENGLPINKNE 1560
DB QTTTFSACPSNSVNOVEDMEIETSEVKKVTSSPITSEESNLNDFIDENGLPINKNE 1560
QY 1561 NVNGESKRKTIVITEVTTMTSTVATESKTVIKVEKGDQKOTVSVSTENCARSTVTTTTVT 1620
DB NVNGESKRKTIVITEVTTMTSTVATESKTVIKVEKGDQKOTVSVSTENCARSTVTTTTVT 1620
QY 1621 KLSTPSTGSGVDIISVKEQSKTVVTTVTDLSLTGGLVTSMTVSKYSTYSDRKVKLMKF 1680
DB KLSTPSTGSGVDIISVKEQSKTVVTTVTDLSLTGGLVTSMTVSKYSTYSDRKVKLMKF 1680
QY 1681 SRPKTRSGTALPSYRKFKVTKKKSIFVLPNDDLLKLARKGGIIEVPYFNNAKPAIDI 1740
DB SRPKTRSGTALPSYRKFKVTKKKSIFVLPNDDLLKLARKGGIIEVPYFNNAKPAIDI 1740
QY 1741 WYPSPRPFTGTTWRYRLQTVKSLAGVSLMLRLLWASLRWDDMAAKVPPGGSGSTRTETSE 1800
DB WYPSPRPFTGTTWRYRLQTVKSLAGVSLMLRLLWASLRWDDMAAKVPPGGSGSTRTETSE 1800
QY 1801 TEITTEITIKRDRVGPYIGIRFEYCIRKICPICPVETPKETPTPQKGLRSLARPKRPE 1860
DB TEITTEITIKRDRVGPYIGIRFEYCIRKICPICPVETPKETPTPQKGLRSLARPKRPE 1860
QY 1861 TPKQGPVITETWABEELELEWEIRAFAPERVEKEKAQAVEQQAQKRLKQOKFTVATSTT 1920
DB TPKQGPVITETWABEELELEWEIRAFAPERVEKEKAQAVEQQAQKRLKQOKFTVATSTT 1920
QY 1921 SPTSSTSTSPISPAQKVMVAPISGSVTTGTGKVLVTTTKVGSFATVTTFOQKNKPHQTEATWVK 1980

QY 781 SFPHNASHRANWIKAVOMCSKPREFALALAILCAVKKPVVMLPIWREELGHTLRHMT 840
DB 718 SFHLNPNWASHRANWIKAVOMCSKPREFALALAILCAVKKPVVMLPIWRESLGHTRLRHMT 777
QY 841 STIEREKEKVKKKKEEETMOQATWVKYTFPVKHQVWKQKGBEYRVTOYGGWSWISK 900
DB 778 STIEREKEKVKKKKEEETMOQATWVKYTFPVKHQVWKQKGBEYRVTOYGGWSWISK 837
QY 901 THYRFPVKLPONTNWNVYKSLGKTKNNMDENMDSDRKRKSRPKKIKIEPDSKDEVK 960
DB 838 THYRFPVKLPONTNWNVYKSLGKTKNNMDENMDSDRKRKSRPKKIKIEPDSKDEVK 897
QY 961 GSDAAGADQNDWISKITEKDDQVKELLDSDSKPCKEPEMEYDDDMKTESHVNCQES 1020
DB 898 GSDAAGADQNDWISKITEKDDQVKELLDSDSKPCKEPEMEYDDDMKTESHVNCQES 957
QY 1021 SQVDVNVVSEGFHLRTSYKKTKSSKLDGLLERRIKQFTLEBKQRLKIKLEGGIKGICK 1080
DB 958 SQVDVNVVSEGFHLRTSYKKTKSSKLDGLLERRIKQFTLEBKQRLKIKLEGGIKGICK 1017
QY 1081 TSTNSSKNLSESPVITKAKGCSQSDMRQEQSPNANNQPEDLIOGCSQSDSVLRMSDP 1140
DB 1018 TSTNSSKNLSESPVITKAKGCSQSDMRQEQSPNANNQPEDLIOGCSQSDSVLRMSDP 1077
QY 1141 SHTTNKLYPKDRYLDVYSIRPETKCPKONSTIENDIEEKVSDLASRGQEPKSKTKGNDF 1200
DB 1078 SHTTNKLYPKDRYLDVYSIRPETKCPKONSTIENDIEEKVSDLASRGQEPKSKTKGNDF 1137
QY 1201 FIDDSKLASADDIGTLICNKKPLIQEESDTIVSSKSAHSSVPKSTNDRDATPLSRAM 1260
DB 1138 FIDDSKLASADDIGTLICNKKPLIQEESDTIVSSKSAHSSVPKSTNDRDATPLSRAM 1197
QY 1261 DFEGKLGCCSESNSTLENSSDTSVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK 1320
DB 1198 DFEGKLGCCSESNSTLENSSDTSVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLK 1257
QY 1321 CELVSGESTGNCEDRLPVKGTANGKKPSQOKKLEERPYNKCSQIKLKNTTDDKNNENR 1380
DB 1258 CELVSGESTGNCEDRLPVKGTANGKKPSQOKKLEERPYNKCSQIKLKNTTDDKNNENR 1317
QY 1381 ESEKKGORTSTFOINGKDNKPKIYKLGECLEKESRVSNGVNPVKNINIKIIPENDIK 1440
DB 1318 ESEKKGORTSTFOINGKDNKPKIYKLGECLEKESRVSNGVNPVKNINIKIIPENDIK 1377
QY 1441 SLTVKESAIRPFINGDVIEMDFERNESSETKSHLLSSDAEGNYRDSLETLPSTKESDST 1500
DB 1378 SLTVKESAIRPFINGDVIEMDFERNESSETKSHLLSSDAEGNYRDSLETLPSTKESDST 1437
QY 1501 QTTTPSASCPSNSVNOVEDMEIETSEVKKVTSSPITSEESNLNDFIDENGLPINKNE 1560
DB 1438 QTTTPSASCPSNSVNOVEDMEIETSEVKKVTSSPITSEESNLNDFIDENGLPINKNE 1497
QY 1561 NVNGESKRKVTIETVMTSTVATESKTIVKVEKDKQTVVSSSTENCASKSTVTTTTVT 1620
DB 1498 NVNGESKRKVTIETVMTSTVATESKTIVKVEKDKQTVVSSSTENCASKSTVTTTTVT 1557
QY 1621 KLSPTSGGSDIISVKEQSKTVVTTVTDLSLTTGGTLVTSMTVSKYSTRDVKLMKF 1680
DB 1558 KLSPTSGGSDIISVKEQSKTVVTTVTDLSLTTGGTLVTSMTVSKYSTRDVKLMKF 1617
QY 1681 SRPKKTRSGALPSYRKFVTKSKTSIFVLPNDLKLARKGGIREVPYFNNAKPALDI 1740
DB 1618 SRPKKTRSGALPSYRKFVTKSKTSIFVLPNDLKLARKGGIREVPYFNNAKPALDI 1677
QY 1741 WPYSPRPTGITHWYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGGSTRTETSE 1800
DB 1678 WPYSPRPTGITHWYRLQTVKSLAGVSLMLRLWLASLRWDDMAAKVPPGGGSTRTETSE 1737
QY 1801 TEITTTTIIKRRDVGPIRFEYCIKRIICPIGVPEPKETPTPQKGLRSSALRPKRPE 1860
DB 1738 TEITTTTIIKRRDVGPIRFEYCIKRIICPIGVPEPKETPTPQKGLRSSALRPKRPE 1797
QY 1861 TPQGTGPVLIETWVAEELELWEIRAFERVEKEKAQAVEQAKKLEQOQKPTVIATSTT 1920

DB 1798 TPQGTGPVLIETWVAEELELWEIRAFERVEKEKAQAVEQAKKLEQOQKPTVIATSTT 1857
QY 1921 SPTSSTTTSTISPAQKVMVAPISGSVTTGKTWLTTKVGSPTVTFQONKNFHOTFATWVK 1980
DB 1858 SPTSSTTTSTISPAQKVMVAPISGSVTTGKTWLTTKVGSPTVTFQONKNFHOTFATWVK 1917
QY 1981 QGQNSGVVQVQOKVILGIIPSSTGTSTQOTFTSFQPRATATVIRPNTSGSGGTTNSQVIT 2040
DB 1918 QGQNSGVVQVQOKVILGIIPSSTGTSTQOTFTSFQPRATATVIRPNTSGSGGTTNSQVIT 1977
QY 2041 GPQIRPGMTVIRTPLOQSTLGKAIIRTPVAVQGAPOOVMTQIIRGQPVSTAVSAPNTVS 2100
DB 1978 GPQIRPGMTVIRTPLOQSTLGKAIIRTPVAVQGAPOOVMTQIIRGQPVSTAVSAPNTVS 2037
QY 2101 STPGQKSLTSATSTSNIOSSASQPPRQOGQVKLTMAQLTQLTQGHGNGOGLTVVIOGQ 2160
DB 2038 STPGQKSLTSATSTSNIOSSASQPPRQOGQVKLTMAQLTQLTQGHGNGOGLTVVIOGQ 2097
QY 2161 QTTGQLOLIPOGVTVLPFGPQOQLMQAAMPNGTVORFLFTPLATTATTASTTTTSTVSTAA 2220
DB 2098 QTTGQLOLIPOGVTVLPFGPQOQLMQAAMPNGTVORFLFTPLATTATTASTTTTSTVSTAA 2157
QY 2221 GTGQROSKLSPQOVHODKTLPPAQSSSVGPAKAQOTAPSAPOQOTOPQSPAQPEV 2280
DB 2158 GTGQROSKLSPQOVHODKTLPPAQSSSVGPAKAQOTAPSAPOQOTOPQSPAQPEV 2217
QY 2281 QTPQEVQOTTTVSSHVPSEAOPTHAQSSKPOVAAQOSQSNVQOSQSVRVQSPQOTRIRP 2340
DB 2218 QTPQEVQOTTTVSSHVPSEAOPTHAQSSKPOVAAQOSQSNVQOSQSVRVQSPQOTRIRP 2277
QY 2341 STPSQLSPGQOSQOVQTTTQSIPIQTPHTSLQIPGOGQPOQPOQVSSQTLLSSQTLNQV 2400
DB 2278 STPSQLSPGQOSQOVQTTTQSIPIQTPHTSLQIPGOGQPOQPOQVSSQTLLSSQTLNQV 2319
QY 2401 SVSSPSRPQOLQIQPOQVIAVPQOLQOQOVVLSQTSQVVAQIOAQOSGVPQIKLOLPI 2460
DB 2320 ----- 2319
QY 2461 QIQQSSAVQTHQIQNVTVVQAASVQEBQORVQOQLRDQOQKKKQOQIENVNTPSKLLIKV 2520
DB 2320 ----- 2319
QY 2521 ELIQOVVKNHNAVTEHLKOKKSMTPABREENORMIVCNVMKYITLDRIDKEEKQAOKKR 2580
DB 2320 -----OVVKNHNAVTEHLKOKKSMTPABREENORMIVCNVMKYITLDRIDKEEKQAOKKR 2374
QY 2581 KREESVEQKRQKQONATKLSALLFKHKEQLRAELIKKRALLDKDLQIEVQEBELKRDLKTK 2640
DB 2375 KREESVEQKRQKQONATKLSALLFKHKEQLRAELIKKRALLDKDLQIEVQEBELKRDLKTK 2434
QY 2641 EKDLMLAQATAVAAAPCPVTPVLPAPPAPPPPPPGVOHTGLLSTPLTPVASQKRKR 2700
DB 2435 EKDLMLAQATAVAAAPCPVTPVLPAPPAPPPPPPPPAVQHTGLLSTPLTPAASQKRKR 2494
QY 2701 EEEKDSSSKSKKKMIISTTSKETKDKTKLYCICTKPYDESK----- 2741
DB 2495 EEEKDSSSKSKKKMIISTTSKETKDKTKLYCICTKPYDESKFYIGCDLCTNWHGECVGI 2554
QY 2742 -----FYIGCDRQCNWHGRCVGIQ 2762
DB 2555 TEKEAKKMDVYICNDCKRAQEGSSBELYICRTPYDESOQFYIGCDRQCNWHGRCVGIQ 2614
QY 2763 SEALIDEVYQCQOSTEDAMTVLPLETEKDEGLKRVLRSLQAHKMAWPFLEPDPNDA 2822
DB 2615 SEALIDEVYQCQOSTEDAMTVLPLETEKDEGLKRVLRSLQAHKMAWPFLEPDPNDA 2674
QY 2823 PDYGVYKPEPMDLATMEERVQRYEKLTEFVADMTKIFDNCRYYNPSSDPFYQCAEVLE 2882
DB 2675 PDYGVYKPEPMDLATMEERVQRYEKLTEFVADMTKIFDNCRYYNPSSDPFYQCAEVLE 2734
QY 2883 SFFVQKLGFKFASRSNNKQSTAS 2907
DB 2907 SFFVQKLGFKFASRSNNKQSTAS 2907

Db 2735 SFFVQKLGFKASRSNNKLQSTAS 2759

RESULT 4
AAB94078 ID AAB94078 standard; Protein; 557 AA.

XX XX
AC AAB94078;
XX XX
XX XX
XX XX
DE 26-JUN-2001 (first entry)
XX XX
XX Human protein sequence SEQ ID NO:14273.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
OS OS
XX EPI074617-A2.
PN PN
XX XX
PD 07-FEB-2001.
XX XX
PF 28-JUL-2000; 2000EP-0116126.
XX XX
XX 29-JUL-0999; 99JP-0248036.
PR 27-AGO-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX XX
PA (HELI-) HELIX RES INST.
XX XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX XX
DR WPI; 2001-318749/34.
XX XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX XX
PS Claim 8; SEQ ID 14273; 2537pp + CD ROM; English.
XX XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX XX
SQ Sequence 557 AA;

Query Match 18.0%; Score 2695.5; DB 22; Length 557;
Best Local Similarity 79.0%; Pred No. 1.2e-129;
Matches 553; Conservative 3; Mismatches 1; Indels 143; Gaps 1;

QY 1937 MVAPISGSVTTGKMLVLTTKVGS PATVTFOQNKSFHOTATFWVKQGSGNSGVQVQOKVL 1996

XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL12838.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Disclosure; SEQ ID NO 32997; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (AB57737-AB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 976 AA;
XX
XX Query Match 9.9%; Score 1488.5; DB 22; Length 976;
XX Best Local Similarity 31.9%; Pred. No. 1.5e-67;
XX Matches 353; Conservative 139; Mismatches 295; Indels 319; Gaps 21;
XX
XX 4 EEEEDGDAETQDSEDEDEMEDDDDDSYPEMEDDDDDASCTESSFRSHSYSS 63
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 99 ESEYHGYDGFDEEDKSDNEDDMLLTPSDDESLEVANESESESVVC---SFNGVCV--- 152
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 64 TPGRRPRVHRPSPILEE-KDIPPLEFPKSSSEDLMPVNEHIMNVIAIYEVLNFGTVLR 122
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 153 ---GRPPRPSPPEVWLQEGRYAALDLPDSEDLEFIANTHRALRSIYEVLRRFRHVR 209
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 123 LSPFRFEDCAALVSQEOCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLYFI 182
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 210 LSPFRFEDCAALACEQSALEHIVHMLLKALLREEDAQGTFFGPDQKQDTVNSLYLI 269
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 183 DGMTWPELVRYCESKEY-----HVLVPOEADDPYGPVENKIKVLQFLVQFLTNIA 238
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 270 DSITWPELVRSYVESDKTDNRNVFHLHSHTS---YPYTGIDNRLEVLPQSLQDFTSNI 326
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 239 REELMSEGVYQDDHCRVCHKGLDLCCTCSAVHLECVKCPLEBVEPDEWOCVCAH 298
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 327 RDVNLQEGPIHYDDHCRVCHRLGDLCCCTCPAVHLECVDPMDNDVPTEDQCGLCRSH 386
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 299 KVPQVTCVAEIQKNKPYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKKIWYSTK 358
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 387 KVSQVDCVLPQEQGLIRHDSLGVDHGRKYWFIAIRIFI-EDQENFT---CWYIST 442
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 359 VQALIEDICLDKDYWEALICKILEEMREETHRHMDITEDLTNKARNSKSFLLAANEETL 418
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 443 SKLLLSRLDAEELTRLHSQITRDERERQMKLTITLNEHKHTKRSVI-----494
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 419 ESIRAKKGDIDNVKSPETEKNKNETENDSKAEKNREPEFQDQSLKSDDKTPDDPQ 478
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 495 -----ETEQAKNELLEKEVLDED--EKDGDAKSESQISIEG 528
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 479 GKSEPTVGDKGNVSANLGDNTNATSEETSPSGRSPVGLSETPDSSNMAEKVAS 538
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 529 TKQEQ-----533
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 539 ELPQDVEEPNKTCESSNTSATTTSIQPNLENSSSSELNSQESAKAADDPENGERS 598
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 534 -----533
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 599 HTPVSIQIEIVGDTSEKSTGELSPGAGKAGSGSTRIITRLNPDLSQLKSOVAA 658
XX | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
XX 534 -----ECKMVTNR-----QKSNQLT- 547

QY 659 AAHEANKLFKEGKEVLVNSQGEISRLSTKEVIMKGINNINNYFKLQGEKGYRYVHNQYST 718
DB | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
548 -----NGLTHFLKMGMEQGFKNYVNOYST 570
QY 719 NSPALNKHQHRDHRKRLAHKFCUTPACGEFKWNGSVHGSKVLTITSLRITOLENNI 778
DB | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
571 NPITALNKPQNEERDKRRHLSHKFSLTASDFKWIGITMTGTTDNNMITTLROTLINEFESNI 530
QY 779 PSSFFHPNASHRANIKAVQCMCKPREFALAILCAVAKPVVWMLPIWREFLGHTRLHR 838
DB | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
631 AASFNLNINWVNNKKNNAVNNARRRSEFAVLLLFQASLKSVVFANVHVEQLGHTTLQR 690
QY 839 MTISIEREKEKVKKKEKKQ---BEETMQOATWVTFPVKHQVWKQKGEETRVTCYGGW 895
DB | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
691 ITSAREERKLEKREKRERDDEERNRLAFNIKYITGLKHQVWKQKGEETRVHGWGW 750
QY 896 SWISKT-----HYRVFPKLPNGTNNYKSLGKTKNNMDENDESDDRKCSRPKKIKI 950
DB | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
751 LMLSSRRRCVRRARQAQLTHNRVYHYTM---GEENDVNEII-----LYDPTQRF 799
QY 951 EPDSEKDEVKGSAAKAGAQ--NEMDISKITEKKDQDVKELLSDSDSKPCKEPMVEVDD 1008
DB | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
800 MQOCSSNVQDQVCHYLPDQYKNVKVIEDVTE-----831
QY 1009 MKTESHVNCQESSQDVVNVVSEGFHL--RTSYKKTKKSSKLDGLLEBRKQFTLEEKQRL 1066
DB | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
832 -KIKGH-----IDVSKALNAPGRTYYSKVARKSRLLDLDLRLRLAEVEQOM-- 877
QY 1067 EKILEGGIKGIGKTSNKKNSSES 1092
DB 878 -ASKIPSDMKPLLVSSQNNTANSKQT 902
XX
XX RESULT 6
XX AAB54364 standard; Protein; 328 AA.
XX ID AAB54364
XX AC AAB54364;
XX DT 09-MAR-2001 (first entry)
XX DE Human pancreatic cancer antigen protein sequence SEQ ID NO:816.
XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
XX KW detection; diagnosis; identification; cytostatic; neuroprotective;
XX KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX OS Homo sapiens.
XX PN WO2000035320-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US05989.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX WPI; 2000-579444/54.
XX N-PSDB; AAC99129.
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
XX treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
XX

XX	DT	18-DEC-2001	(first entry)	
XX	DE	Novel human secreted protein #2435.		
XX	KW	Human; vaccination; gene therapy; nutritional supplement;		
XX	KW	stem cell proliferation; haematopoiesis; nerve tissue regeneration;		
XX	KW	immune suppression; immune stimulation; anti-inflammatory; leukaemia.		
XX	OS	Homo sapiens.		
XX	PN	WO200179449-A2.		
XX	PD	25-OCT-2001.		
XX	PF	16-APR-2001; 2001WO-US08656.		
XX	PR	18-APR-2000; 2000US-0552929.		
XX	PR	26-JAN-2001; 2001US-0770160.		
XX	PA	(HYSE-) HYSEQ INC.		
XX	PI	Tang YT, Liu C, Drmanac RT;		
XX	PI	WPI; 2001-611725/70.		
XX	DR	Nucleic acids encoding a range of human polypeptides, useful in genetic		
XX	PT	vaccination, testing and therapy -		
XX	PS	Claim 20; Page 529-530; 765pp; English.		
XX	CC	The invention relates to novel human secreted polypeptides. The		
XX	CC	polypeptides and antibodies to the polypeptides are useful for		
XX	CC	determining the presence of or predisposition to a disease associated		
XX	CC	with altered levels of polypeptide. The polypeptides are also useful for		
XX	CC	identifying agents (agonists and antagonists) that bind to them. Cells		
XX	CC	expressing the proteins are useful for identifying a therapeutic agent		
XX	CC	for use in treatment of a pathology related to aberrant expression or		
XX	CC	physiological interactions of the polypeptide. Vectors comprising		
XX	CC	the nucleic acids encoding the polypeptides and cells genetically		
XX	CC	engineered to express them are also useful for producing the proteins.		
XX	CC	The proteins are useful in genetic vaccination, testing and		
XX	CC	therapy, and can be used as nutritional supplements. They may be used to		
XX	CC	increase stem cell proliferation; to regulate haematopoiesis; and in		
XX	CC	bone, cartilage, tendon and/or nerve tissue growth or regeneration;		
XX	CC	immune suppression and/or stimulation; as anti-inflammatory agents; and		
XX	CC	in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid		
XX	CC	sequences of novel human secreted proteins of the invention.		
XX	SQ	Sequence 246 AA;		
Query Match 5.2%; Score 781; DB 22; Length 246;				
Best Local Similarity 67.8%; Pred. No. 3.3e-32;				
Matches 162; Conservative 10; Mismatches 63; Indels 4; Gaps 2;				
Qy	877	HQVWKQGEYRYVTGGWSWISKTHYRFPVKLPNTNVRKSLGKTKNNNDENDES	936	
Db	7	HQVWKQGEYRYVTGGWSWISKTHYRFPVKLPNTNVRKSLGKTKNNNDENDES	66	
Qy	937	DKRCSRPKKIKTEPDSEKDEKVGSDAAKADQENMDISKITEKKDQVKKLLSDSDK	996	
Db	67	DKRCSRPKKIKTEPDSEKDEKVGSDAAKADQENMDISKITEKKDQVKKLLSDSDK	126	
Qy	997	PCKEPMEVDMDKTESHVNCBESSQVD--VNVVSEGFHLRTSYKKTKSKLDGLLERR	1054	
Db	127	PCKERTN--GSRKHENRVTKLSGFEFSRCGCGXXGFSKDXLQENKIIQTRWTSXRE	184	
Qy	1055	IKQTLTEKORLEKIKLEGKIGKIGTSTNSSNLSESPVITKAKGCGSDSMRQSQP	1113	
Db	185	NXTVYTGKTRATRNQVGGWNGXGYRKDSTNSSNLSESPVITKAKGCGSDSDETRETP	243	
RESULT 12				
ABB62887				
XX	ID	ABB62887 standard; Protein; 997 AA.		
XX	AC	ABB62887;		
XX	DT	26-MAR-2002 (first entry)		
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO 15453.		
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;		
XX	KW	pharmaceutical.		
XX	OS	Drosophila melanogaster.		
XX	PN	WO200171042-A2.		
XX	PD	27-SEP-2001.		
XX	PF	23-MAR-2001; 2001WO-US09231.		
XX	PR	11-JUL-2000; 2000US-191637P.		
XX	PR	23-MAR-2000; 2000US-0614150.		
XX	PA	(PEKE) PE CORP NY.		
XX	PI	Venter JC., Adams M, Li PWD, Myers EW;		
XX	PI	WPI; 2001-656860/75.		
XX	DR	N-PSDB; ABL06990.		
XX	PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
XX	PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
XX	PS	interactions -		
XX	PS	Disclosure; SEQ ID NO 15453; 21pp + Sequence Listing; English.		
XX	CC	The invention relates to an isolated nucleic acid detection reagent		
XX	CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
XX	CC	useful in developmental biology and in elucidating cell signalling and		
XX	CC	cell-cell interactions in higher eukaryotes for the development of		
XX	CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
XX	CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
XX	CC	sequences (ABL01840-ABL16175) and the encoded proteins		
XX	CC	(ABE57737-ABE72072).		
XX	CC	The sequence data for this patent did not form part of the printed		
XX	CC	specification, but was obtained in electronic format directly from WIPO		
XX	CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX	SQ	Sequence 997 AA;		
Query Match 4.4%; Score 661.5; DB 22; Length 997;				
Best Local Similarity 24.7%; Pred. No. 3e-25;				
Matches 309; Conservative 157; Mismatches 395; Indels 389; Gaps 53;				
Qy	1520	DMEIETSEVKKVTSSPITSEESNL-----SNDPID---ENGLPINK	1558	
Db	6	NVEIESTSIVTGLNSGNAEDVDMPGRRRRKQKSKSYIGTKDVLDTLDKDIPLNK	65	
Qy	1559	NENVNGESKRKTVITEVTTMTSTVATESKTVIKVEKGDQVWVSSTENCASKSTVTTTTT	1618	
Db	66	-----QNRFPIT-----APVUREC-----VKYERETFEENRV-----	97	
Qy	1619	VTKLSTPTGGSV-----DIISVKEQSKTVVTTTDTSTLTGGTLVTSMTVSKEYSTRDK	1674	
Db	98	---YSTSPRGVRLNDAAKLYEQ-----VKTEDK	126	
Qy	1675	VKLMEFSPKKTPTSTALPTSKFTKTSIFVLNDDLLKLRKSGIREVPYFNVA	1734	
Db	127	STITK--KPSYSR----YPLTSNFTLHKRSLVLPREFELKLARLGKSGSTNGFHAA	180	
Qy	1735	KPALDIWVPSPRPTFGITWRYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGGST	1794	
Db	181	KNN-TIWQCSRPLFRFCWSTYSNATSLSSALQLRLWASLRWDDMAAKVPPGGST	238	

Db 384 -----IDLPSSTEEKSESTAAEEDIVKVTTPPEGGSEEDVPKPSQIPEKIT 434
QY 443 ETE-----NDSDAKNREE-----FEDQ-----SLEKSDDKTPDDPEQG 479
Db 435 EDELLKVTSAPAKASPEEEVVATTATSAPTEEDYKPTTAGTISEEBEGKPTP---AEEG 491
QY 480 KSEE-----PTEVGDKGNSVANSGLDNTVNTATSEETSFSRGRS-----P 518
Db 492 SGEEEKDVKVTAAPEETEDEAKPTSAPY---ASDEKEQEPKPSGSGDELDLKTPTAPT 548
QY 519 VCCLSETPSSNWKAKKVASLELPQDVPB-EPNKTCESSENTSATTTSIQPNLENSNSSEL 577
Db 549 AGATGASESEBQDGGK-STEAPTSVDIDIEPAKPTSESEBASG-----EGEDVAKET 599
QY 578 NSQSSESAAADDPENGERSHTPVSI-OEEIV-GDFTSEKSTGELSSPAGKAGSST 635
Db 600 TPAGEASIAAGEEIVKGTTPAGEPSSEGEDEIVKGTTPAESSESEDE----- 648
QY 636 RIITRLRPDSKLSQLKSOQVAAAHAANKLFKEKE--VLVNSOGESIRLSTKKEVIM 693
Db 649 --LTKVTTTPAGEPSVAGEEIEAKETTPAGEPSIAGEEIVKVTTPAGE-SSIAGEEIVK 705
QY 694 KGNINNYFKLGOEGKRVYVHOYSTNSFALNKHQHREHDHRRHLAHLKFCLTTPAGEFKWN 753
Db 706 -----VTPPAGESSESSEGEIEIK-----VTPPAGESSE 734
QY 754 GSVHGSKVLT-----LSTRLITOLENNIPSSFFHPNWSHRANKIKA--VOMCSKPRE 806
Db 735 GDEEIVKESTPAGEPSEGEEDVIKATTSAPKSDIEGVKEPETATEVPAEEVEDFAKP-- 792
QY 807 FALALALECAVPVVMPLPWREFLGHLRLHRMTSIEREEKKVKKKKEKKEEETMOQA 866
Db 793 ---TTPAEAEPIAGTPIPTDGIS-----GEEIVKGTTPOTLEQEPSEERS 839
QY 867 TWKYTFPPKHQVWKOGEEYRVGYGGWSWISKTHVRFVKPLPGNTVNW-YRKSLECT 925
Db 840 TEV-----PVAED-----DLSSTSSASAIASSTEGV 865
QY 926 KNNNDENMDSDRKRCSRPKIKIBPDSEKD-----EVKGSAAKAGADQN 971
Db 866 ODAASETTSAPARAGDKDEAATVTAQDKDDVEEQDATDLPVEDVQVSTAKTTTTE 925
QY 972 EMDISKITEKQDVKELSDSDKPCKEPMEVDDDMK----- 1010
Db 926 QPKBESSTEADEAEIEVTTSSPADKQEVPEAPADKHDEEDVQATDLPKISDTCPPV 985
QY 1011 --TESHVNCQESSQDVVNVYSEGHIL-RTSYKKTSSKLDGLLERRIKOFTL-----E 1061
Db 986 VDEATTGQPETSETATDKPPSVYILPPVSOEVPSSSTAKVNDNRNDETETKPTLPPSGEDQ 1045
QY 1062 EKQLEKIKLEGKIGK-----IGKTSNKKSLSESPVITKAQECQSDSMRQOESPNA 1115
Db 1046 SSEPLPAMDLPAGIPGEGDCLVEGKTVANNITVPAUTAPCDVSK--CISLVAQCOQMECK 1103
QY 1116 NNDQPE-----DLIOQCSOS---DSSVLRMSDPSHTNKLKPDVRLDDVSIRSPTKC 1166
Db 1104 LPENLEKCTVAADLLDGCCTPIYICDESTESAEDDEESTAK--PDNKIDEDVSEISTE-EI 1160
QY 1167 PKQNSIENDIEEK-----VSDLASRGOEPTKSKTKGNDFIDDSKLASADDI 1213
Db 1161 PKDVIMPTGITEQPLSHVKPDEEIOPTVSVPAQFDESTTAKVKKP--IDES----- 1210
QY 1214 GTLICKNKKPLIOBESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMDFGKLGCDSESN 1273
Db 1211 ---AEDKKPIGESEED-----SKPIDSEEDKKPVE----- 1238
QY 1274 STLENSDTVSIQDSSEDMIV-----QNSNESISEQFTREOD-----VEVLEPLKE 1322
Db 1239 ---ESAEDKRVEDSEEEKPLPTVIPASEIEKSEKPEDEKKTADFAATEQPEATPA 1295
QY 1323 LVSGSTGNCBDRLPVKVTEANGK---KPSQOKLEERPVNKCSDQI---KLKNTTDKN 1376

1296 QIADTAEKVEYDDKLATTAPVSGEDELKPADEKKRTE-----TAQIPDAEIPASTDEPE 1349
1377 NENR---ESEKKGQRTST--FGINGKDNKPKIYLKGECLKEISESRVSVGNVPEKVNNI 1430
1350 SSTEPLTVUDLKPEEDSKGTAPESDKVPVPTASTENEIEESDKFTTVPAPKISAS 1409
1431 NKIIP--ENDIKSLTVKE-----SAIRPFINGDVI-----MEDFFERNSSSEKSHL 1474
1410 DETEPTAEEDLVPAFTPEIESEFEVSTKKPAVOGPPPLPTLAPAOPEKKPVDAETSTEADI 1469
1475 LSSSDAAGNVRDSLELPLSTKESDSTQTTTP-SASCPESNSVNOVDEMETEITSEVAKVTS 1533
1470 STEPSAEVEKESGETSESNEIDAGASSTPVPVPADEDKTPTSTKVE---ADDFKFTV 1526
1534 SPIT-SEESN---LSNDFIDENGLPI-----NKNENVNGSK-----RKT 1570
1527 APLAGDEESNLKLPQIDIFEEEA-PVAVTTAAAPSKDD---GEQKPVVEEKEPIEDQGP 1582
1571 VITEVTTMTSVATESKTVIKVEKGDQKQTVVSSSTENCAKSTVTTTITTKLSTPSTGGS 1630
1583 IEDETSTPTS-----SENEIEPESDRATTAPSKEE--PSEPSTGAPTKDEPAEPSTDP 1635
1631 VDILSVKEQSKTVVTTTVDLSTTGTGLVTSMTVSKKEYSTRDKVKLKMFRSKPKTKRSGT 1690
1636 ESDS-KETPESEVPTIVAPA---GEKIPTSSITPDEEPTATAPVAKPDEDEKETST 1690
1691 ALPSYRKFTVKSTKKSIFVLPNDDLKLRKGGIREVPYNYNAKPAIDWIYPSPPRTF 1750
1691 EIPTDAPASSEDENS---STDQIPS-----EVP-----EKKP----- 1720
1751 GITWYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGG---GSTRTETSEITETTE 1807
1721 -----ETAQTPESGDIVGATAAPTSDSEVPVQ 1749
1808 IIKRDRVGPYRIRFEYCIRKILCPIGVPEPKETPTPQRK-GLRSSALRKPAPETPKQTG 1866
1750 RLPE-----EVLAEIPQPSSTETGIRKQOEDTAAPSIDAK-- 1783
1867 PVIIETWAELELEWEIRAFARVEK--EKAQAVEQ-----OAKRLEQOKPVIATST 1919
1784 ---EPYVTEIDEEATVAPISEKDEKTEEEKPVEQKPTGEEPESEEEKEKPIEQDVT 1839
1920 TSPTSTSTTISPAQKVMVAPISGVTGTGKMLVLTTKVGSATVTFQONKFNHOFATWV 1979
1840 EGPVTEASEAGSTES-----SEEVKPSTEGEVAEKPED----- 1873
1980 KQGQNS-GVVOVQOKVILGIIPSSTG---TSQOFTSPQRTATVIR-PNTSGSGGTS 2034
1874 KOPSSTAQAPVETIPEISTELPAQDGDKPTSEAPVDSDEDSAPSDEKIPVSG----- 1927
2035 NSQVITGPIRPGMTVIRTPLOQSTLGRKAIIRTPVNVQFGAPQV-MTOIIRGQPVSTAV 2093
1928 --EEVEGPEV-----TTASPOAAEDELKTPAESSEPSSTDKVPETEKPEDEKTRAD 1977
2094 SAPNTVSSPPGKSLTSTST-----SNIQASASOPPRPOQOQVKLTWAQLTQLTQGHGG 2148
1978 ETPESTVQSDVATSTAPVAGVAGDIEKDEQATTASP--EEEEIEKPTIAPAEIPQ----- 2031
2149 NOGLTVVIOGQFTTQOLQIPQGVTVLP--GPGQOLMO-AAMPNGTVQRFVLTPLATTA 2205
2032 -----PSEKEPVDE-OEVESGKATPAESDGGQPIDEIAATSGPIDE-----ASTA 2076
2206 TTASTTTTIVSTTAAGTGEQRQSKLSPOMQVHQDK-----TLPPAQSSSVGPAKAQPT 2259
2077 APTKEESTIVASAA-----SP--AVHDEIKDVTTPQVDAKEVAAPQDETKT 2123
2260 A-----QPSAR-PQOTQOPQSPAQPEVOTQEVQT---QTTVSSHVPB-----EAQPTH- 2304
2124 SIDVSTDSPTAQDDEKQDKTEAPVAPTIVSSSTADSAASSTPTVEPVEIDKPMDD 2183
2305 --AQSSKQVA-AQSQPQSNVQOSQPVV--QSPQSTRIRPS-----TPSOLSPQOOSQ 2353
2184 IMSQTIAPHTADGAASTSTDEDEDAQPVTVSPQDAEKTPTVSPAPQSDSKTSPSEAPQDADE 2243

2564 RNVKHSFGARNSTRIASDSQSQAPADLTIKOEHPAIAQAEIKRQLSDSEQKFK 2623
1052 ERRIKQFTLEKQBLEKIKLEGGIKGKTSTNNKSLSESPVITKAKECQSDSMRQEB 1111
2624 SRHDSNSSIER---KLKTEREI---KTELGDYNSSEYTVTGKLKEYSPETKRKKHK 2675
1112 SPNANQDPELLOGCSQSDS---VLRMSDPS---HTNKLYPKDRVLDVSI-- 1159
2676 SKRR-----LKSSSTADTSAAQTPLVMTPLPSPFDVHSSECKTKFEDDLKTEC 2727
1160 -----RSPETKCPKONSIEIIEEKVSDLASRQEQPTKSKTKGNDFIDDSK 1206
2728 SSIPLEISAGERKHKEKREKLNMTEATVPN-----SPTNDT-----SSE 2773
1207 LASADDIGTLCKNKKPLIQEESDTIVSSSKSALHSSVPKSTNDRDAPLSR----- 1258
2774 KLSKEERHLKSKSKSMDSNCTKIYNSSGA-H--PSTSPSLPATPTTAPSTAQTSK 2829
1259 ----AMDFEGLGCDSESNSTLENSDTSIQDSEEDMIVONNESISEOFRTREQDVE 1314
2830 RGEDKMEIFGLISDEESQPEQETNKKDILIPS-----VSTTGPIVSAALQYKQ--- 2881
1315 VLEPLKCELVGESCNCEDRLPVKGTANGK---KPSQOKKLEERPVNKCSDQIKLN 1370
2882 --EP-----STPNS-----KNEBAHIQLTVHEPEQOQOLER---SRLSGSSSSS 2921
1371 TTDKKNENRESEK---KGQRTSTFOINGDKNPKIYKLGCELKEISESRVSVGNVEPK 1426
2922 HADRHRREKREKREKRSQEQOQNIHOKSSK-----VETK 2959
1427 VNNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFERNESSEYKSHLLSSDAEGNYR 1486
2960 VDDNSVMDDEAGRALEAQ-----LMSDFDKPISE----- 2990
1487 SLETLPTSK---ESDSTOTPTSPACSPNSVNOV-----EDMEIETSEVKVTSPPIT 1537
2991 --EATESTAATYRSDMTDFRESDN-EDNNSVDMTKQGVKSEQBPQHSKDKKKKKRKS 3047
1538 SEESNLNDFIDENGLPINKNENGVESKRTVITEVTMTSTVATESKTIVKEKDGK 1597
3048 EEKQELL-----QOORRESLPNVASTSSAPPTPKLTVNQAAASK 3088
1598 QT-----VVSSTENCAKSTV-----TTTTTIVKLTSTPGSDVLIIVSKQS 1640
3089 HADLQDKAHISPPVCKPSPSLPCLIGDDDDALHTPAKPTTPSSRGNDGLTSPREK 3148
1641 KTV-----VTTVTVDLSLT-----TTGGLTVTS----- 1662
3149 RLISPIPKPTTIANSTLTQSAETPVSSGTVISSALATTTSTTAAGVSAAPGLDNP 3208
1663 -----MTVSKYST-----RDKVKLMKTSRPKK 1685
3209 TSASAOCKKKEFIPGFDGLDDRESAVOSISAEFNSTSLDNDIADEPKIPVASPPRA 3268
1686 TRSGTALPSYKRVKTVKSTK-----SI 1707
3269 TKPLDKLEESKSRVTISQEETSEAVSALLGESFGTSTTDDYSLDGMDEMSVNELETPTL 3328
1708 FVLPNDDLLKLRK-----GGIREVPVFNNAKPAIDWPYSP-----RPTFGIT 1753
3329 VIAEPDEEALAKAIETAGEPASILEEP-----EMEPERAEAPDPPEAEISEPVVEVL 3384
1754 WRVRL-QTVKSLAGVSLMLRLLWASLRWDDMAAKVPPGGSGTSTETSETTEITTEIKRR 1812
3385 DPELNKAVOSLKHEDMM-----DIKADTPQSERDLQIDT-DTEENPDE---AD 3429
1813 DVGYPGRFECYKIKIICPGVETPKETTPQKRLRSALRPKRPETPKQGVPIIET 1872
3430 SSGP-----SKIDETVQSSSSSEKSSINNS-----PIPRETANDI-- 3466
1873 WVAEELELWEIRAFARVEKEKAQAVEQOAKRLEQOKP-----TVIATSTSTST 1926

3467 -----PNVESQPKLSNESTPOPSVITKLPFLDTPKTPAGLPPSPVKIE 3510
1927 TSTISPAQKVMVAPISGSV---TTGTRKMLVTKVGSPTATVTFQONKNEHOTFATWVKOG 1982
3511 PPTISKLOQPLVQVQTVLPAPHSSTGSGISANSVINLDSLNSVSSCNTSAASATASASA 3570
1983 QNSGVVVOQKVLGIIPSTSTGTSQOTFTSFOP-RTATVTRPNTSGSGGTSTN---SQV 2038
3571 SISFGSPTASQANAM---POASTPKQGPITPOQAIQTOSLIMQPPFTISIEQTPHFAVQOM 3627
2039 ITGQPI---RQC--MTVIRTPLOQSTL---GKAIIRT-----PVMVQPGAQO 2078
3628 VLSPOSHHQPQPTQYVMGIRAFSPHPLSPGSGVAQSLVQOLSPVGRPMYSQP-SPQO 3686
2079 VMTQIIRGQPVSTAVAPNTVSTPGOKSLTSATSTNIQSSA---SQPPRQOQOVKLT 2135
3687 --QVQOTQOQHALLITSPQSSNISPLASPTTRVLSSNSPTTSKVNYSQPRNQ----- 3737
2136 MAQLTQLTQGHGNGOGLTVVIQOGQGTGOLQILPQGVTVLPQPGQOLMOAAMPNGTVQR 2195
3738 -----VPQSPSPKSAEVAEQTPQLMTI---PQKMTPIQVPHIPT-- 3774
2196 FLFTPLATTATTATTTTSTTAAGTGEQORQSKLSPOMQVHODKTLPPAQSSSVGPAKA 2255
3775 -IISKVTVQPOQATQSOVASSPPLGS-----LPPHKNVH-----LNAHQN-----QO 3816
2256 QPOT-AQPSARPOPTQPSPAQPEVQTPQEVQOTQTVSSHVSPSEAPTHAOSKSPQVAA 2314
3817 QPQVIKMTAHOHQHMQOQFMHQMTQROQHMO-----QOQLHGQSQ--QIYIS 3862
2315 QSQPOSNVQSGSPVRVQSPQSTRPSTPSQLSPGQ-QSOVQTTTQSPQIPQIPHTSLQIP 2373
3863 APQOHMQHQAQOQOQHQQHNLHQAQHTQKQHQAOQOQFNQ-QIQQHQSQO-- 3918
2374 SQGQPSQPQVQS---STQTLSSGQTLNVQSVSPSPRQIQIOQOPQ-QVIAVPOLQOQ- 2428
3919 -QHQVQOQNAQOQHLQSOQHQSQOQLNQ---QHOAQOQOQLOQIQKLOOHGPGQOQKS 3973
2429 -----VOVLSQISQOVVAOIQAOQSGVPOQIKLQLPQIQOQSSAVOTHQIONY 2476
3974 PQGVGHGSGTSIFASQOHNSQLPAR-----GVPOQ---QHPQQLSHSSPCRPNTLVSV 4024
2477 -VTVOAASVQQLQVQQLRQOQKQKQO-----QIEINV--NTP 2513
4025 NQGVQPPALLRVGSHSQPNQOQLPHQSSSGHPHQKOLSSPGANLPLQTPLVNIQNT 4084
2514 SKLLIKVEII-OKV---VMKHNAVIEHLKQKSKMTAEREENORMIVQNMVYILDKI 2569
4085 -KIIVQOHIVAQNVPPPTQGNAIHYPNQNGKDSPTPGH----- 4123
2570 DKEEQAAKRRKREESVEQKSKQNAKLSAL-----LPKHKEQLRAEILKRALLDK 2622
4124 -VEPTPMSAOKTSESVSVITPTTGLAVISANTVGSLLTEENLIKISQPKQDELIEQ 4182
2623 DLQIEVQ-----EELKRDLLKIKK-----EKDLMOLAQATAVAAPCPVTPVLPAPAP 2670
4183 DSK-EVDSYNSAKEVNDISVIKKLDTPLASKDAKRAVEMQAI-----AP 4226
2671 PPSPPPPGVQHTGLLSTPLPVASQKREEEKDSSSKSKKKM 2715
4227 APIPNQPGNQ--SMAQETALPTTSMVNSNNDHDETEDETETROL 4269

RESULT 15

ABB71160

ID ABB71160 standard; Protein; 5560 AA.

XX

AC ABB71160;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 40272.

XX

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
 KW Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX N-PSDB; ABL15263.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions.
 XX Disclosure; SEQ ID NO 40272; 2lpp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 5560 AA;
 SQ
 Query Match 4.1%; Score 613.5; DB 22; Length 5560;
 Best Local Similarity 19.5%; Pred. No. 9.6e-22;
 Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;
 QY 305 DCVAEQKNKPYIRHEPIGDRGRKRYWFLNR-----RLIIEEDT 344
 DB 1912 DTADKAEKQ--RHEKEKKEQEKREKDLRKQVEREKRKAQAEEREKEDKAKEEK 1968
 QY 345 ENENEKKIYKTVQLAEIDCLDYWEAEICKILEEMREIHRHMOITE-DLTNKA 403
 DB 1969 EREREKKAQEDREK-----KEREERELRE--KEORKEQKEKEIREKDLREKEQ 2015
 QY 404 GSNKSFLLAAANEIILESIRAKGDIENVKSPETEEDKNTENDSKDAEKNREFEQSL 463
 DB 2016 -----REDRNREKELDKDUREKEMREKEREKELHREKQDREHR 2057
 QY 464 EKDDDKTDPPDPE--QKSEEPTE-----VGDKGNSVANLGDNTTNATSEETSPS 513
 DB 2058 EKEQSRAMDVQEGRGMRRELSYQKSKMDIAGEASSITADCQHNKENAMDTIAQGT 2117
 QY 514 EGRSPVGLSETPDSNMAEKKVASELPQDVPEPNK---TCESNNTSA-----TTTSI 564
 DB 2118 PGASP-----STP--SDNTPKERSKRLSRNSPYRLHRRLLSQESNHSAGCGGSGSSH 2170
 QY 565 QPNLENSNSELNSQESAKAADDPENGERS--HTPVSIQEEIVGDTSEKTSIGELS 622
 DB 2171 QIHEDYVVRIRMENSONSVSHSNQRLNDRNDREKHEKSSFKED-----KNSSHIS 2223
 QY 623 ESPGAGKAGSGSTRIITRLRNPDSKLSQKSOQVAAAHAANKLFREKGEVLVYNSQGEI 682
 DB 2224 RPHGCGGSSASSSKHHR-----RDKHHQKGSASSIETNSSI-----EYVVDPISQYK 2271

QY 683 SRLSTKKEVIMKGINNNYFKLQEGKRYRVYHNOYSTNSFALNK-----HOHR----- 729
 DB 2272 HNLNTSEELQSHQ-----PKREKEREHFSSHANSSRRHKSKRDHHRKKRHSVA 2324
 QY 730 -----EDHDKRRHLAHKFCLT---PAGEFKWNGSVHSGKVLTIISLRTITIQLENNIP 779
 DB 2325 ESTNTDEEHTPOQHNRHRIISAAGSGSAGELSSAANTSGKLHHQHHRKSRKSRGS 2384
 QY 780 SSFFHPNASHRANWIKAVQMCSPREFALALALECAVPVVMPLPIWREFLGHTRLHRM 839
 DB 2385 DEGHSSKSLRAKMLMLSSADSDDTDA-----SKKHSI 2419
 QY 840 TSIERE-----EKEKVKKKEKQEBEETMQOATVVKYFPVKQVWVKQGEYRVTYGG 894
 DB 2420 FDIPTDCPNVSMYDKVKARCKNNQRAEKKIKAKFS-----QLKQSAKKRKRSTYDG 2474
 QY 895 WS---WISKTHVYRFVPKLPNTNVNRYKSELECTKNMDEMDESCKRCSRPKKTKIE 951
 DB 2475 DSDTEFEDROH-----RNSGSSFFHGR--YPLGSSDDDDDEETHORRIS----- 2517
 QY 952 PDSEKDEVKGSAAKADQNE-----MDISKITEKKDQDVKELLDSDSKPCKEEPMEDV 1006
 DB 2518 -----SDSAEHGGODNQGAFTLADANVRQMO-QNLRRLCDGDS--SED--EIR 2563
 QY 1007 DMKTESHVNCQESSQDVVNVVSEG-----FHLRTSY-----KKTSKSLDGLL 1051
 DB 2564 RNVKSHFGKRNSTRIASDESQSQAPDLTIKQEHPIAQAQIKREQLSDEQKFK 2623
 QY 1052 ERIKQFTLEKORLEKIKLEGIGKIGTKTSTNSSKNLSESPVITAKEGCQSDSMRQEQ 1111
 DB 2624 SRHDSNSEEER---KLKTEREI---KTELGDYFNSSEYTYTGKLUKEYSPETRRKHKH 2675
 QY 1112 SPANANDQPEDLIQGCQSQSS-----VLRMSDPS---HTTNKLYPKDRVLDVDSI-- 1159
 DB 2676 SKRR-----LKSSSTADTSAQPLVMTPLTPTSFVHSSECKTKEDFNDDLKTEC 2727
 QY 1160 -----RSPETKCPKONSTENDIEEKVDSLASRGQPTKSTKGNDFIDDSK 1206
 DB 2728 SSIPLEISAGERKHKERKREKLRNMTATVPN-----SPTTNDT-----SSE 2773
 QY 1207 LASADDITGLICKNKKPLQIEESDITVSSSKSALHSSVPKSTNDRDATPLSR----- 1258
 DB 2774 KLSKEHRHLKSKSKSDNSCNTKIYNSGA-H---PSTPSLPATPTPSAPSTAQTSK 2829
 QY 1259 ----AMDFEGLKCDSESNSTLSESDTVSIQDSSEEDMIVQNSNESISQFRTREQDVE 1314
 DB 2830 RGEDKMEFIFGIISDEESQFPEQAETNKDIIPSS-----VSTTGPIVSAALQTYKQ-- 2881
 QY 1315 VLPLKCELVSGESTGNCEDRLPVKTEANGK-----KPSQOKLEERPVKNCSDQIKLN 1370
 DB 2882 --EP-----STPNS-----KNEEAHIQLTVEHEPQQOQLER---SRLSGGSSSS 2921
 QY 1371 TTDKKNNENRESEK---KGORTSTFOINGKDNKPKIYLKGECLKEISESRVSVGNVEPK 1426
 DB 2922 HADREHRRKREKKREKRSQREQQNQIHOXSK-----VETK 2959
 QY 1427 VNNKIIPENDIKSLTVKESAIRPFINGVIMEDFNERNSSSETKSHLLSSDAEGNYRD 1486
 DB 2960 VDDNSVDMDEAGRALEAQ-----LMSDETKPISE----- 2990
 QY 1487 SLETLSTK---ESDSTQTTTPSASCPSNSVNOV-----EDMEITSEVKKVTSSPIT 1537
 DB 2991 --EATPTAATYRSDMTDVERFSDN-EDNNSVDMTKQGVSEQEQEKKDKKKKKRKS 3047
 QY 1538 SEESNLSNDFIDENGLPIKNENNVNCESKRKVITEVTMTSTVATESKTVIKVEKGDK 1597
 DB 3048 EEKQEKLL-----QQORRESLPNVASTSSAPPTPGKLTNVNQASK 3088
 QY 1598 QT-----VVSSTENCAKSTV-----TTTTTIVTKLSTPTGGSDVIISYEQS 1640
 DB 3089 HADLQLDAKHISPPKCPSPSLPCLIGDDDDDALHTPKAKPTTPSSRGNDGLTPSREKP 3148
 QY 1641 KTV-----VTTVTVDSLT-----TTGTGLTVS----- 1662

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 00:43:52 ; Search time 30.6646 Seconds
(without alignments)
9116.791 Million cell updates/sec

Title: US-09-698-295-1

Perfect score:

Sequence: 1 MVSEEEEDGDAEETQDSE.....KLKGFKASRSHNKLQSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs. 96168682 residues

Total number of hits satisfying chosen parameters: 2833308

Minimum DB seq length: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%
Listing first 45 summaries

Database : PTR 76.4

Database : FIK_6: 1: pirl:

```
1: pir1:
2: pir2:
```

```

3: pir3:
3: pir2:
3: pir1:

```

5: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %			DB	ID	Description
		Match	Length				
1	4082	27.3	810	2	G01252	small GTP binding	
2	854.5	5.7	1711	2	T21432	hypothetical prote	
3	546.5	3.7	5327	2	T13564	microtubule-associ	
4	514.5	3.4	2447	2	T16870	hypothetical prote	
5	490	3.3	5170	2	T15348	hypothetical prote	
6	463.5	3.1	3507	2	T34513	hypothetical prote	
7	462.5	3.1	3759	2	A35085	trithorax protein	
8	458.5	3.1	3488	2	T34418	hypothetical prote	
9	446	3.0	2897	2	B48666	cell proliferation	
10	440	2.9	2271	2	F90073	hypothetical prote	
11	439.5	2.9	3256	2	A48666	cell proliferation	
12	439.5	2.9	3924	2	S37431	ankyrin 2, neuronal	
13	439	2.9	2938	2	T30249	cell proliferation	
14	419.5	2.8	2232	2	T34434	hypothetical prote	
15	417	2.8	2441	2	D71623	erythrocyte membra	
16	407.5	2.7	3828	2	T13857	trithorax protein	
17	406	2.7	4377	2	A55575	ankyrin 3, long sp	
18	405.5	2.7	452	2	T21435	hypothetical prote	
19	403	2.7	5105	2	T32650	hypothetical prote	
20	402.5	2.7	510	2	T21430	hypothetical prote	
21	401.5	2.7	405	2	T21433	hypothetical prote	
22	401.5	2.7	1367	1	S48478	glucan 1,4-alpha-g	
23	397	2.7	2722	2	T20532	hypothetical prote	
24	397	2.7	6642	2	T29757	protein UNC-89 - C	
25	393	2.6	1630	2	A53577	ascites stialoglyco	
26	393	2.6	2481	2	D90011	FmtB protein [impo	
27	392	2.6	1871	2	D96796	probable heat shock	
28	392	2.6	2738	2	E88320	protein F07A11.6 [
29	391	2.6	1939	2	T18372	repeat organellar	

RESULT 1

G01252

small GTP binding protein SEC4 homolog - human

C;Species: Homo sapiens--(man)

C:\projects\rome\api.htm

C>Date:	21-Dec-1996	#sequence	revision	06-Jun-1997	#text change	29-Jan-1999

C:Accession:-G01252-

R; Bowser, R. P.

submitted to the EMBL Data Library, January 1994

A;Reference number: G06441

A;Accession: G01252

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-810 <BOW>

A;Cross-references: EMBL:U05237; NID:g1276427; PID:g451848

C;Genetics:

A;Gene: FAC1

Query Match	27.3%;	Score 4082;	DB 2;	Length 810;
Best Local Similarity	98.73%	Pred. No. 3.2e-144;		
Matches	777;	Conservative	4;	Mismatches 2;
			Indels	Gaps
			1;	
QY	1	MYSEEEEDGDAEETQDSEDDDEDEMEEDDDDDSDYPEEMEDDDDDASCTESSFRSHST	60	
Db	1	MYSEEEEDGDAEETQDSEDDDEDEMEEDDDDDSDYPEEMEDDDDDASCTESSFRSHST	60	
QY	61	YSTPCRRPRVHRPSPPILEEKDIPPLPEPPKSSDDLMPNENHIMVIAIYEVLRNFGTV	120	
Db	61	YSTPCRRPRVHRPSPPILEEKDIPPLPEPPKSSDDLMPNENHIMVIAIYEVLRNFGTV	120	
QY	121	LRLSPRFEDFCAALYSQEQCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLY	180	
Db	121	LRLSPRFEDFCAALYSQEQCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLY	180	
QY	181	FDGWTWPEVLRYVCSDEKEYHHVLVYQAEADYPYGPVENKIKVLQFLVDQFLTNNIARE	240	
Db	181	FDGWTWPEVLRYVCSDEKEYHHVLVYQAEADYPYGPVENKIKVLQFLVDQFLTNNIARE	240	
QY	241	ELMSEGVIOYDDHCRVCHKLGLDLCCECTCSAVVHLBCVKPPLLEEVPDEMQCEVCVAHKV	300	
Db	241	ELMSEGVIOYDDHCRVCHKLGLDLCCECTCSAVVHLBCVKPPLLEEVPDEMQCEVCVAHKV	300	
QY	301	PGVTDCAEIQKNKPYIRHEPTGYDRSRKRYFNLNRLLIEEDTENENEKKIWIYKTKVQ	360	
Db	301	PGVTDCAEIQKNKPYIRHEPTGYDRSRKRYFNLNRLLIEEDTENENEKKIWIYKTKVQ	360	
QY	361	LAELIDCLDKDWEALCKILEEMREEIHRHMDITEDLTNKARGSNKSLAAANEILES	420	
Db	361	LAELIDCLDKDWEALCKILEEMREEIHRHMDITEDLTNKARGSNKSLAAANEILES	420	
QY	421	IRAKKGDIDNVKSPEETEKDKNETENDSKDAEKNREEFEDQSLSEKSDDKTDPDDPEQOK	480	
Db	421	IRAKKGDIDNVKSPEETEKDKNETENDSKDAEKNREEFEDQSLSEKSDDKTDPDDPEQOK	480	

QY	481	SEEPTEYDVGKNSVSNALGNTTNTATSEETSPSEGRSPVGCSETPDDSSNMAEKKVASL	540	QY	325	DRSRKYWFLNRRLLIEDTENENKKIWIYSTVKVQLAELIDCLDKDYWEAELCKILEEM	384
Db	481	SEEPTEYDVGKNSVSNALGNTTNTATSEETSPSEGRSPVGCSETPDDSSNMAEKKVASL	540	Db	490	DRYGRYWFIVRRLVWOSLDETE---LYYSTVPOLYQLQKLDRTTYEYKDLCDTIRLR	545
QY	541	QDVPEPNKTCSSNTSATTTIQPNLNSNSSSSELNSQSSESAAADDPNGERESHT	600	QY	385	REEIHRHMDITEDITNKARGSNKSFLLAAANEIILESIRAK--GDDNVKSPSETEKDK	441
Db	541	QDVPEPNKTCSSNTSATTTIQPNLNSNSSSSELNSQSSESAAADDPNGERESHT	600	Db	546	IDFLQMALVEMTSERR-----BAALETWVKROLIGYDAEATTPQ-----	588
QY	601	PVSTQOEIVGDTSEKSTGELSPGAGKAGSGSTRIITRLRNPDSKLSQLKSOQVAAAA	660	QY	442	NETENDSKDAEKNREEFEDQSLKSDSDKTPD---DDPEQCKSEEPTEVGDGKNSVSNL	498
Db	601	PVSTQOEIVGDTSEKSTGELSPGAGKAGSGSTRIITRLRNPDSKLSQLKSOQVAAAA	660	Db	589	-----IYLHRDSMKRMASILRDCAQKGOVKQEVKL-----	618
QY	661	HEANKLFKEGKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNOYSTNS	720	QY	499	GDNNTNATSEETSPSEGRSPVGCSETPDDSSNMAEKKVASLPODVPEEPNKTCESSNTS	558
Db	661	HEANKLFKEGKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNOYSTNS	720	Db	619	-----EE--PVEGQSPVKCVQFVEDSI-----LPE-----	641
QY	721	FALNKHQHRDHKRRHLAHLKFCITPAGEFKWNGSVHGSKVLTITSLRLTITQLENNIPS	780	QY	559	ATTTISIQPNLNSNSSSSELNSQSSESAAADDPNGERESHTPVSIQDEIVGDTSEKST	618
Db	721	FALNKHQHRDHKRRHLAHLKFCITPAGEFKWNGSVHGSKVLTITSLRLTITQLENNIPS	780	Db	642	-----SMGIFDAKLINTFWSGGA-----TOEELVEQFV---DI	672
QY	781	SFFHPNW 787		QY	619	GELSESPGAGKAGSGSTRIITRLRNPDSKLSQLKSOQVAAAAHEANKLFKEGKEVLVNS	678
Db	777	TSLHPSF 783		Db	673	SDNEDAPSA-----	681
RESULT 2				QY	679	QGEISRLSTKKEVIMKGNINNYFKLGQEGK---YEVYHNOYSTNSFALNKHQHRDHDKR	735
T21432				Db	682	-----NLWRMGDEGNDQTFMTYTYNSRNMDSSEFTLRKAADKK	721
hypothetical protein F26H11.2 - Caenorhabditis elegans				QY	736	RHLAHLKFCITPAGEFKW---NGSVHGSKVLTITSLRLTITQLENNIPSPFFHPNASHR	791
C:Species: Caenorhabditis elegans				Db	722	KYASKF--AQIDNEDVWVAKNRQFYGDASLHCKFIMTLOQVINKNIPIDLHMRKWPEFA	779
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999				QY	792	ANWIKAVQCMQPRFALALALECAVPVVMPLPWREFLGHTRLHRMTSITEREKEKVK	851
C:Accession: T21432				Db	780	KGFDLEVSADDDYKKLVTCLLKLDCAVRKTIKMPQWNGLGOTLERITVVDRENFMKEQ	839
R:Barlow, K.				QY	852	KKEKK-----QEEETMQQATWYKTYFPVKHVKWOKGEYRVTG---YGGSWISKT	901
A:Reference number: 219421				Db	840	QRLKIDADALYKDLDDSFVRVYNYMKWPNTY--ILRQGETYRNAGKSGGAWAAK	898
A:Accession: T21432				QY	902	HVYRFVPLKPGNTNVNYRKSLEGTKNMNDENDESCKRCKSPKKIKIEDPSEKDEKVG	961
A:Molecule type: DNA				Db	899	YVEKWI-----QVPESPKPLPLAVTVEIKTESVSNRK-----	930
A:Residues: 1-1711 <WIL>				QY	962	SDAAKADQNEQMDISKITEKKDQDKVLELSDSDKCPKEEPMEVDDDKMTESHVNCQESS	1021
A:Cross-references: EMBL:Z81515; PIDN: CAB04197.1; GSPDB: GN00020; CESP: F26H11.2				Db	931	-----ARLELVSKIT-----	942
A:Experimental source: clone F26H11				QY	1022	QVDVNVVSGFHLRTSYKKKTSKSLDGLLERRIKQFTLEEKQRLKIKLEGKIGIKGT	1081
C:Genetics:				Db	943	-----KKQRSG-----GKS	952
A:Gene: CESP:F26H11.2				QY	1082	STNSSKNLSESPVITKAKGQSDSMRQEQSPNANNQDQEDLIQCSOSDSSVLRLMSDPS	1141
A:Map position: 2				Db	953	SKKPTPELT-----NGCYSPPSCRS-----	971
A:Introns: 56/3; 154/2; 373/2; 417/3; 505/2; 596/1; 703/2; 847/3; 922/3; 1034/2; 1078/2;				QY	1142	HTTNKLYPRDRLVDDVSIRSBETKCPKQNSIENDIEKVVDLASRGOEPTKTKGNDFF	1201
Query Match				Db	972	-----NPNKRC-----YSP-----	980
Best Local Similarity				QY	1202	IDDSKLASADDIGTLICKNKKPLIOBESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMD	1261
Matches				Db	981	-----MCRN-----	984
4				QY	1262	FEGKLGCDSESNTLSSSDTSVSIQDSSEEDMIVQNSNESISFQFRTREDDVEVLEPLKC	1321
Score 854.5; DB 2; Length 1711;				Db	985	-----	984
Pred. No. 3.5e-24;				QY	1322	ELVSGESTGNCEDRLPVKGTGTEANGKKPSQOKKLEERPVNKCSQDKLKNNTTDDKKNENRE	1381
Mismatches 296; Gaps 72;				Db	985	-----GVLVSQAKQAHDERKLEE-----	1001
Indels 1139; Gaps 72;							
Conservative 296; Mismatches 624; Indels 1139; Gaps 72;							
QY	4	EEEEEE---DGADEETQDSDE---EDEMEEDDDSDYPPEMEDD---DDDASYCTE	52	QY	4	EEEEEE---DGADEETQDSDE---EDEMEEDDDSDYPPEMEDD---DDDASYCTE	52
Db	169	KROEEDIIYMDSESEESSESSDDEFMLNEDQVQEEELNLTIDIKIEKGLDENKYC--	226	Db	169	KROEEDIIYMDSESEESSESSDDEFMLNEDQVQEEELNLTIDIKIEKGLDENKYC--	226
QY	53	SSFRSHSYSTPGRRKRVHRPRSPILKE--KDIPLPFPKSSSEDLWPNNEHIMVIAI	110	QY	53	SSFRSHSYSTPGRRKRVHRPRSPILKE--KDIPLPFPKSSSEDLWPNNEHIMVIAI	110
Db	227	-----PWLDEPASPPLKLEPSESSQDIPITASIMDAVEI	261	Db	227	-----PWLDEPASPPLKLEPSESSQDIPITASIMDAVEI	261
QY	111	YEVLRNFGTVLRSLPFRFEDFCAALVSOEQCTLMAEMHVLLKAVLREEDTSNTTFPAD	170	QY	111	YEVLRNFGTVLRSLPFRFEDFCAALVSOEQCTLMAEMHVLLKAVLREEDTSNTTFPAD	170
Db	262	YELRSYHRTLRITPTFEDFCAALISHNNSCIMAEVHMALLRNCLKSDDEQTHYSVTE	321	Db	262	YELRSYHRTLRITPTFEDFCAALISHNNSCIMAEVHMALLRNCLKSDDEQTHYSVTE	321
QY	171	LKDSVNSTLYFDGTWTFEVLRYVCESKDYHVLVPOEA-----EDYPYGVENKI	222	QY	171	LKDSVNSTLYFDGTWTFEVLRYVCESKDYHVLVPOEA-----EDYPYGVENKI	222
Db	322	TNNSNIMIHMDTLTYAEILRQYIEA-----YPFADASVRDAINVDNTPFVGYDANI	374	Db	322	TNNSNIMIHMDTLTYAEILRQYIEA-----YPFADASVRDAINVDNTPFVGYDANI	374
QY	223	K-----VLQFLVDOFLTTNIAREBELSESGVIOYDDHCRVCHK--LGDLLC	265	QY	223	K-----VLQFLVDOFLTTNIAREBELSESGVIOYDDHCRVCHK--LGDLLC	265
Db	375	QRDFSEFFNKKHVFRLVLLFWMSYFLLYSSEFKLVNNVGVKQFQDENCRCVGKSSGRVVG	434	Db	375	QRDFSEFFNKKHVFRLVLLFWMSYFLLYSSEFKLVNNVGVKQFQDENCRCVGKSSGRVVG	434
QY	266	CETCSAVYHLECKVPLBEVPEDEWQCEVCAHK--VPGVTCVABEIQKNKPIYRIHEPIGY	324	QY	266	CETCSAVYHLECKVPLBEVPEDEWQCEVCAHK--VPGVTCVABEIQKNKPIYRIHEPIGY	324
Db	435	CTQCEAAAFHVEC--SHLKPFPE--VLVCNICKKNSAVRGVLPPPEDAVDREP--LRSQPIGR	489	Db	435	CTQCEAAAFHVEC--SHLKPFPE--VLVCNICKKNSAVRGVLPPPEDAVDREP--LRSQPIGR	489

```
Qy 1382 SEKKQRTSTFQINGKONKPKYILKGCLKEISESRVSVGNVPEKVNINKIIPENDIKS 1441
Db 1002 -----
Qy 1442 LTVKESAIRPFINGDIVMEDFERNSETSKSHLLSSDAEGNYRDSLETLPSTKESDSTQ 1501
Db 1002 -----
Qy 1502 TTTSPASCPESNVQVEDMEIETSEVKVKTSSPITSEESNLNDFIDENGLPINKNEN 1561
Db 1002 -----SG 1003
Qy 1562 VNGESKKKTVITEVTTMTSTVATESKVIVKEGDKQTVVSVSTENCAKSTVTTTTTIVTK 1621
Db 1004 VLGEK-----
Qy 1622 LSTPSTGSDIISVKEQSKTVTTTDTSLTTTGGTLVTSMTVSKVSTRDKVKLMKFS 1681
Db 1010 -----
Qy 1682 RPKTRSGTALPSYRKFTVKTSKSIKIFVLPNDLKLARKGGIREV--PYFNYNKAPALD 1739
Db 1010 -----AWPIPEIQETSTKRGKSIKIFVQKILROMIMGCGCOQVMPGFSAGIKSNLL 1062
Qy 1740 IMPYSPRPFTGITWRYELQTVKSLAGVSLMLRLWASLWDDMAAKVPPGGGSTRTETS 1799
Db 1063 IMPYAPRPRLDLCWKQWTLNARSLHAVALQKLIWSSIKFNEFD---PDDTHPDRRWI 1119
Qy 1800 ET--EITTEILKRDVGPYRFEYCIRKIICPI-GVPTPKETPTPQRKG-----LR 1850
Db 1120 DTSHDERRIRHKEMPPGQYERYEMEIEIPLYDEPEEDESWLSRNGSSSFHSR 1179
Qy 1851 SSALRPKRPT-----PKQT-----GPVITWVAEELELWE----- 1883
Db 1180 SSSARKRPORHEFLSLKFGNPKKNAFRSLDNRRATAIRREWVDGVLKVFEEKDYW 1239
Qy 1884 --TRAFARVEKEKAQV-----EQAKRLEQOQPTVIATSTTSSTSTSTISPAQ 1934
Db 1240 KWIHAEAKTAKRLEATRAKAKAKEDERRRQQOQOORSVARIPV-PMHS-----LIPSE 1294
Qy 1935 KVMVAPISGSV---TTGTKMLVTKVGSPTATVFQO-----KNFHOTFATWVKQO 1983
Db 1295 RNNV-PYLGSOQORRPNNGERFLEKYNSSSVSPQAHGYASTPPGCHQPPQNIIRQAG 1353
Qy 1984 SNSGVVQOQKVLGIIPSSGTSGOQTTSPQPTATVIRPNTSGSGGTTSSNQVITGPQ 2043
Db 1354 YNQ-----LPRKPTTSPNFQS-RP-VATITPTPOLRAAAGADG---VVRAMV 1396
Qy 2044 IRPG---MTVIRTPLOQSTLGRKAIIRTPVMQVQAPQOQVMTQIIRQO----- 2087
Db 1397 MTPGNKSTVNTNSTPYQO-----ALNRQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1451
Qy 2088 -----PVSTAVSAPNTVSTPGQKSLTSAT-----STSNTOSS 2120
Db 1452 NPSVQMHQRLPONRAALQRFPGESTTMRVRTEAAIPDNDGDEQPPVPIPRYDPTSNFDAQ 1511
Qy 2121 ASQPPRQOQOQOQKLTMAQLQOLQGHGNGOGLTVVIOGQGTGQLOL----IPQGVTVL 2176
Db 1512 RAQOQHQPQRPVSTPAQMIRTTQPGGVKH--NVILMKASDGTOKMVLKPGQPPGTVI- 1568
Qy 2177 PGQOQLMQAAMPNGTVORELFTPLATATATTTTTVSTTAAGTGEORQSKL----- 2230
Db 1569 -STGQVVVVRQPTAVOQROLYT-----ATPGTRVVIRPNANGAPROQDHQVRRVVQ 1621
Qy 2231 --SPQMVHODK--TLPPAQ-----SSSVGPAKAQPOQAQPSARPQOPOQOS 2274
Db 1622 ASGPRAMEYDDOGTPPGQOVRYVLOGGSGTIPNPNPKVSSRGGPRGLTMQWVQOQO 1681
Qy 2275 PAQPEVQTOPEVQT---QTTVSSHVPSAQ 2301
Db 1682 QHNPAYHMDPDDATGFAVSTTTEQVPDEQO 1711
```

RESULT 3

Tl3564

microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: Tl3564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: Tl3564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.7%; Score 546.5; DB 2; Length 5327;
Best Local Similarity 18.3%; Pred. No. 3.9e-12;
Matches 550; Conservative 479; Mismatches 1134; Indels 825; Gaps 121;
Qy 2 VSEEEEDGDAAETQDS-----EDDEE-----DEMEEDDDSDYPPEMEDDD 44
Db 2191 VKDEAQESRRSKTESVPPKAKDDKSPKEVLQPVSMETITREDADQPMKPSQASRR 2250
Qy 45 DD-ASYCTESFRSH-----STYSTPGR-----RKPRV-----HRPSPILEE 82
Db 2251 ESTAESIKASPRDEKSPKASKEASRFGSVAESIKYLDLKPQIIKDDKSTHSRESLED 2310
Qy 83 KDIPPLEFPKS--SEDLAVPNEHIMNVIAIYEVLNFGTVLRSFERPED-----FCAALV 136
Db 2311 KSAVTE--KSVRPLSVASDH--PAVAIEDDAKS-----SISP---KDKSRPGFVAETV 2359
Qy 137 SQ--EOCTLMAEMHVLLKAVLR-----EEDTSNTFGPADLKDSVNSTLYFID 183
Db 2360 SSPTTEATMEFSKTEVVEKSSLSLQGGSGGKQTDSSPDVDAEGDFSHVASVSTVTP 2419
Qy 184 GMTWP-EVLRVYCESDKYEHVLPYQAEADYPGP-----VENKIKVLQ 226
Db 2420 TLTKPAELAQIGAAKTVS---SPLDEALRTPSAPEHISRADSPAECASEIASQDKSPQ 2475
Qy 227 FLVDOFLTNIARELASEGVIOYDHCRCVKLGLDLCCTCSAVYHLECVK----- 280
Db 2476 VLKSSRPAAVASEK-----DDAAQLKSSVEDLR-----SPVASTEISRASAGET 2521
Qy 281 ---PLEEVPDEMOCEVCAHKVPGVTDCAVIEIQKNKPYIRHEPIGYDRSRKKYWFNLNR 337
Db 2522 ASSPIEAPKDFAEFE--QAEK--AVLPLTIELKGNLP----- 2555
Qy 338 LIIEEDTENENEKKIWTYSYTKVQLAELIDCLDKDYWEALCKLIEEMREEIHRHMDITED 397
Db 2556 -----TLSSPDVVAHASVQPAELSKV-----DIET 2581
Qy 398 LTNKARGSNKSFALAAANEETLESIRAKKGDIDNVKSPSETEKDKNETENDSKDAEKNEE 457
Db 2582 ASSPIDCAPKSLIGSPAE-----RPESPAESAKDAEASEVSKSDASRPPSV 2628
Qy 458 FEQSLKEDSD-----KTPDDDPQOQK--SEETPEVGDKGNSVANI--GDNNTNAT 506
Db 2629 VESTKADSTKGDISPSPESVLEGPDKDDVEKSKSSRPP-----SVSASITGDKTQDS 2681
Qy 507 -----SEETSPSEGRSPVCLSETPDSSNMNAEKKVASELPQDVPEPNKTCSSN 556
Db 2682 RPASVWESVKDEHDKAESRRRESIAKVESVIDEAGSKSKSS---QDSQDKSTLASKE 2738
Qy 557 TSATTTTIQPNLENSSSSELNSSQESAKAADDPGERSHTPVSIQE-----EIVGD 611
Db 2739 ASRESVY----ESSKDDAEKSESRPESVIASGEPV--PRESKSPLDKSTSRGSPVES 2792

Db 4579 EEPADVDERSKVESKSIATIMTSIIRKPSDEM--EPISK--LVEEHE--HVELAQ 4633
QY 2634 RDLKIKKEDLMQALQATVAAPCPVPVYL-----PAPPAPPPS 2673
Db 4634 EVTSTKTTLLQSSQSSTTSSTKTGASRESITLTQMDQQTQSOGDPADKTPPT 4693
QY 2674 PPPPGVQ 2681
Db 4694 APVSPGVK 4701

RESULT 4
T16870
hypothetical protein T13H2.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000
C;Accession: T16870
R;Wu, X.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid T13H2.
A;Reference number: Z18593
A;Accession: T16870
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2447 <NUX>
A;Map position: X
A;Cross-references: EMBL:U39653; NID:g1049397; PID:g1049400; PIDN:AAB52494.1; GSPDB:GN00
A;Experimental source: strain Bristol N2; clone T13H2
C;Genetics:
F;158-207/Domain: RING finger homology <RRN>

Query Match 3.4%; Score 514.5; DB 2; Length 2447;
Best Local Similarity 18.2%; Pred. No. 2.2e-11;
Matches 514; Conservative 423; Mismatches 1037; Indels 857; Gaps 118;

QY 9 EDGAEETQSDDEDEMEEDDDSDYPEEMEDDDDDASCYTSS----- 54
Db 2 DDSPGPGTSKARDKAENABENTSDSSDSEVSASEKSESRPSSEKKKVIITRVIPRP 61
QY 55 -----FRSISTYSTPGRRKPRVHRPSRILEKEIPPLEFPKSSDLMPVNEHIMVI 108
Db 62 PTRDKGHRVNLLESNESETKSLYQRAKEGIPSYKGKPEIKLPTTSEQYYDLEVLN-- 119
QY 109 AIYEVLNFTVLRSPERFEDFCAALVSQEOCTLMAEMHVLLKAVLREEDT'SNTFGP 168
Db 120 -----PARMEG-----RELTNA----- 132
QY 169 ADLKDSYNSLYFIDGNTWPEVLRYCESDKYHHVLPYQEAEDYPYGPVENKIKVLQFL 228
Db 133 ---YDAVRNKNYVLP-----KSVCEAD-----LQKV 156
QY 229 VDQFLTNTIARELMSEGVYQYDDH--CRVCHKGLDLLCETCSAVYHLCEVAPLLEVP 286
Db 157 IGSF-SCDVQELIQSIIMTKGHRFCQOCILVAEWRSGNTCPTCRQLNLSKRELQODP 215
QY 287 E-DEMOCEVCVAHKVPGVTDVCAEIOKN-----KPIYRHEPIGYDRSRKRYWFLNRR 337
Db 216 RFDLIYQVVSISVG--RWMAENREHEKDVYFGRKYIEG--GSDMNKRYGIDPNSK 270
QY 338 LIIEEDTENENKKI-WYSTKVQLAEILDCLODYWEALCKILEMREIEIHRHMDITE 396
Db 271 LKAPRLKSAGRKKIRWFHES-----DED--GSVRKVMESKGA-----PKE 310
QY 397 DLTNKARGSNKSLAANEILSIRAKGIDID---NVKSPETEKDKNFTENDSKDAEK 453
Db 311 DFTNYLNDKEGTSVAAEKEVLE-----EGEMDFPIEKSSDEQTDLDDEESMLDSDF 365
QY 454 NREEFEDQSL-----EKSDDKTFDDDPGCKSEETEVGDKGNSVSANL 498

Db 366 EISDNEDEVKPSQSTSKKTTNRSRDSSESNDSDRNELQKKRKM-----KRKNVPKTD 419
QY 499 GDMTNTATSETPSEGRSPVGCCLSETPSSNMAEKVASELPQDVPPEPNKTCESNTS 558
Db 420 GSDVSNESFEDASGE-----VVATKL-----IKESKKKPCGRPKK 456
QY 559 ATTTSIQPNLENSSSELNSSQSESAAAD-----DPENGERSHETPVSIQE 606
Db 457 FAPELIEGDIPTPSEDSLTSSDEERDDNAADPAVFQKBFNRDP---RRDGH----- 506
QY 607 EIVGDFTSKSTGELSSEPGAGKAGSGSTRIITLRNPDSKLSQLKQQVAAAAHEANKL 666
Db 507 -----PEKD-----KLYNDFMIDM-----NHQVDRK 528
QY 667 FKEGKEVLVYNSQGEISRLSTKKEVIMKGINNYFKLQEGKYRVYHNOYSTNSFALNKH 726
Db 529 FEKXGEITHVISDDNSNSESDEAE-----DRESSIDS----- 560
QY 727 QHREDHDKR--RHLAHKFCLTPAGEFKWGSVHGSKVLTTSTLTLTTQLENNIPSPFFH 784
Db 561 -----EHEKEISKFLSHRQPLNPNTSVDDDCQV-----ITVVKKD----- 595
QY 785 PNWASHRANWIKAVQMCSPREFALALALECAVKPVVMLPIMREFLGHTRLHRMTSIE 844
Db 596 -----VKOSATSKP-----GETSPDSSSKIE- 617
QY 845 EEKEVKKKKKEKQEEETMQOATWVYTFPVKQHVQKQGEYRVTYGGWSWISKTHVY 904
Db 618 EKPDKV--SEEVSDDEMTPEHITADKGTDTFLNIMEHDDM-----YGG-----YLF 663
QY 905 RFVPKLPNGTNNVYRKSLETKNNMDENMDESCKRCSRPKIKIIEPD-----SEKDEVK 960
Db 664 R-----PGDGTGIS--RPKVQAPGNTNLSMNVCPVAVLKEGKKLVIPDEYIESSDET 717
QY 961 GSDAAGAQONEMDISKITEKKDQDVKELSDSKPCKEEPMVDVDDDMKTESH----- 1014
Db 718 LSDSEETSPAEQMEQSETSEAGSTIITKSGTERETOGSSPSEPTSRKMKHKLDT 777
QY 1015 -----VNCQESSQVYVNVSEGFHLRT-----SYKKTKSKSLDGLERRIKQ 1057
Db 778 RRRKLADSDLSDFDVSIDGNELVATGPKPIIKHKVFDYDANRMPKSNLDFTRRRNARE 837
QY 1058 FTLEEKORL-----EKILEGIGIKGTSTNSKNLSESPVITAKGCCQSDSRQOE 1110
Db 838 IPMEETSRLAEEQVAHEEYKIHRRRQVYVLEAVEAASKLVNYVDVTTEEE-----IEEE 891
QY 1111 QSPNANNQDIEDLIQCSQSDSVLRMSDPSHTT---NKLYPKDRVLDVDSIRSPETKC 1166
Db 892 ETPB-----EYVK--VASPTAPIATENPTTSTAPFEEGVAMKETPIEIFF--DPDEPC 942
QY 1167 PKNSIENDI-----EEKVSDLASRGOEPTKSKTKGNDFFIDDSKLASADDIGTLIC 1218
Db 943 SSAQAQARELIIERVGKEQIIEDSLQNRKPS-SKT-----VKESREAA----- 987
QY 1219 KKKPLIQESDRIVSSSKALHSSVPKSNDRDPTPLSRAMDFEGKLGCDSESNSTLEN 1278
Db 988 --QEPRIKDEMESEQKDADNPTVEVDKESASESSDKSDFE-----DETLD 1036
QY 1279 SSTVTSQDSEEDMIVQNSNESISEQFTR-----EQDVEVLEPLKCELVGESGNC 1332
Db 1037 QSKTVKLSLKHETV-----SDEI-EDFDTKFEFVATADAKMKIRTIIGETVSTEFLL--- 1088
QY 1333 EDRLPVKTEANGKPSQKKLEERP-----VNKSDQIKLNTTKKNNENRESEK--- 1384
Db 1089 -----KLVAQQPAVTVDEVLAGFCVRNTDQEFSTIKETCKRTKNKPNDDSMV 1135
QY 1385 KGORTSTFQNGK-----DNKPKYIKGCELKEISESRVSGNVKPPKNNIKIIPEN 1437
Db 1136 KNFRES-FAAKHRPVPRKPLTNTERMYI--ERAHMVKYKHV--DMEPLHMKILIALQKQ 1190
QY 1438 DIKSLTVKES--AIRPFINGDVIMEDFNERNSETKSHLLSSSDAEGNYRDSLETLPST 1494
Db 1191 QIAATCANUSQPVTVTPPEEHAEQVQLLHNLONPSILRPLL-----NNPQFALTLLHKAQ 1243

Db	1002	QVTSETTTTTVTREYNDEPEDEHQPHSPAPSSHTAEHPHIVETTTTTVTREFOE	1061
QY	483	EPEVEV-----GDKGNSVANLGD-----TNNATSETPSE--GRSPVGCSTPDS	528
Db	1062	EPELEYKQEDNSRKSPSSHQSQENLVETTTTTVTSEYDPEPEFEQETGKSSPAPS	1121
QY	529	SNAEKKVASLPODVPEPNKTCSSNTSATTSTIQPN-LSNSN-----SSELNSS	580
Db	1122	HVESERQV---LESVPASDPRHVMTETTTTTVTTRQPHDDSESRDSPNRDVVEAQSIHSS	1178
QY	581	QSESAAADDPENGERSHTPVSIOEIV--GDFTEKSTGELSESPGAGKASGSTRII	638
Db	1179	NIESHOQFSEKDDSGSPVSKSEDEPVKHESYKFETSTTE-----TREF	1224
QY	639	TRLRNPDSLKLSQKSOQVAAAHAANKLEKEGKVLVWNSQGEISRLSTKKVIM-KGNI	697
Db	1225	DR-----PELDSERLSEPA-----QSPMEVSPITTEENIVRPSVV	1260
QY	698	NNFKLGQBGKY-RVYHNOYSTNSPALNKHQHREDHKRRHLAHKFCLPAGEFKWNGSV	756
Db	1261	KSEY--GSEGHVPSVJETTTTTVTREFYDDQDEHENQTSPELASSIPTBEEHEGS-	1317
QY	757	HGSKVLTITLRLTITOLENNIPSPF-----EHPNASHRANWIKI-----VOMCS	802
Db	1318	HLFKETTTTT--TVTREFYDEPNVEELODPQFSPAPSSHVESGHHASESPVAQOQETP	1374
QY	803	KPREF-----PHEEYESHVLTEQAPLLTEQHQHPESGDSDGEGLSKVLFACK	819
Db	1375	QTRFHEFSPAAYPHEEYESHVLTEQAPLLTEQHQHPESGDSDGEGLSKVLFACK	1434
QY	820	P-----VVMPLWREFLG-----HTRLHRMTSIE-REEKEKVKKKKEEETMOQATW	868
Db	1435	AGVAGVGVAAAPVALAAVAKAAYDALKKDDDEEDQEEERESLLRQERSIDSPHASEQS-	1492
QY	869	VKYTPPVKHQVKQK--GEERYVTCYGWMSWKTHVYFVP-----KLPG--NTWNYR	919
Db	1493	QIEEHERFEESVPSEKHHTVTTTTVTREFYDFEHEPLVSQEIEGEGKSPADSE	1549
QY	920	KSL-----BCKNNMDNDESKRCKSRPKKIKLEPDSEKDEKVGSDAAKAD---	970
Db	1550	KSLPHVETTTTTVTREFDKNDSE--SPVPSEKQEPITVSREV--YETABGEPEHHY	1605
QY	971	NEMDISKITEKDDQVVELLDSDSKPKCKEPEV-----DDMKTE---	1012
Db	1606	TETTTTTVTKEVIDDSQEM--GDDDEPKQESQVVTETTTTTMTSREYDNDDETSEAGD	1663
QY	1013	SHV-----NCOESSQVDV-----NVSEGFHLRTSYKKKTKSS	1045
Db	1664	SHITETKTTVVVREFHGGQPEETEETDEVELPPKIEEDNVSEYSESSTSVSREVRPD	1723
QY	1046	KLDGLLERRI-----KQFTLEEKQRLKTK	1070
Db	1724	E-PHIIETTTTTVTREYHNEPEEYDDQDAAPISFSQEHQDDDSQAASHDQHDRESPE	1782
QY	1071	LEGGIKGICKT-----NSKNLSSSPVITKAKE--GCQSDSMRQEQSPNA-	1115
Db	1783	SEKSVKHTTETTTTTVTROLYDDEASEIRGESPVATEEHEHVSVSTKSDDESEHQHVPVI	1842
QY	1116	-----NNDQPEDLIQCSQSDSSVLRMSDP-----SH-----TTN	1145
Db	1843	EYTTTTVTREFYDDQDE--LOREDHTQSEERSSIPTETEHEHDSHLIKETTTTTVTTR	1900
QY	1146	KLYPKDRVLDDVSIRSPETCKPKQNSIENDIEBEKVSDLASRGOEPTKSKTKGNDDFIDS	1205
Db	1901	EYDEPENVE--KLQDSQFSLSPSSHVESEIYVPESPVAKQEIPTRE-----PHDS	1952
QY	1206	KLAS---ADDITGLCKNKPLIQEESDTIVSSKSALHSSVPKSTNDRDATPL-SRAMD	1261
Db	1953	PAAYFHEDEYEHQVTEQAPLLTEQ-----HQ--PESGEESDGEFGSKVLG	1999
QY	1262	FEGLKGC-----DSENSTL-----ENSDTVSIOD	1287
Db			
Db	2000	FAKKAGMWAGVVVAAAPVALAAVAKAAYDALKKEDDEDEOQERESLIREERSDFDPSHASE	2059
QY	1288	SSEEDMIVONSNESISEQFTRREQDVVEVLEPLKCELVSQES-TGNCE-----	1333
Db	2060	QSOIEKEHRRFEESPVPSEKHHDDSSALPOESVSQPIEXESRTFNDESEFGVKSDDHYTED	2119
QY	1334	DLPLVKGTTANOKKPSQ--OKKLEER---PVNKCSDQI-KLKNWT---DKNNENRESEK	1384
Db	2120	DOESLKSPPSEGEAFQFTSEKQDRSDSPIHOSKEDIQFQNESPEVDVKSQPHDEEK	2179
QY	1385	KG-QRTSTQINGKONKPKIYLKGECLKEISESRVVSNGVPEKVNINIIIPNDIKSLT	1443
Db	2180	PDLEROGSSSGSYSPKSP-----GGSITGLDEEKALSGVQEPEDR-----PENFAESHE	2228
QY	1444	VKESAIRPFTINGDVIIMEDFERNSSSTKSHL-----LSSSDAEGNYRDSLETLPSTKES--	1497
Db	2229	KTEAT-----SDENLFESDKYAPASPVPSESDSSNRVETTTTTVTREHFE	2274
QY	1498	-----DSTQTTTPSACSPESNVQVDEMEIETSEVKVKVTSPTISEEP--SNLSN	1546
Db	2275	PEDDHSYVVESOEYSYSSSGSPVPSEKSDRV---IETTTTTVTREHFEADEIPTIVES	2330
QY	1547	DFIDENGLPINKNENNVGESKRTVITEVTMTSTVATE-----SKTVLKVBKGDK	1597
Db	2331	SHDDQAASVPSEEDVHEQ-----IOTTTTTVTKEHFVDPDDDEIDSEHMNESDKYAS	2383
QY	1598	QTVVSTENCACAKSTVTTTTVTVK-----LSTPSTGGSDII---SVKEQSK	1641
Db	2384	GSPVPSEEDSSRVETTTTTVTREHFEPEDDHSPVVOQYQSASESPVPSEKSERVIE	2443
QY	1642	TVVTTTTVT-----DSLTTTGGTLVTSMTVSKE-----YSTRDKVKLKMFP	1680
Db	2444	TTTTTTVTREHFEDEHILQCOGESDQIPSEISITSENMDRETSSSPVSNRDEEFVLP	2503
QY	1681	SRPKK-----TRSGTALPSYRKFTVKTKKSI FVLPNDDLKLARKGGTREVYPFN	1731
Db	2504	IAPYKOPTYGRVDSHDADPASA-----PSAESPIQAYKQESQAHSLEQEQQRSSVS	2557
QY	1732	YNAKPAALDW-----PYSPRPTFGITWRYRLQTVKSLAGVSLMLRL-----	1774
Db	2558	HEDSPAAYFHDNESDHDSPVSDRAPLLTEQHQHPESGESDGEFGSKVLGFAKAGM	2617
QY	1775	-----WASLRWD-----DMAAKVPPGGGSTR-----TETSETE	1802
Db	2618	VAGGVVAAAPVALAAVAKAAYDALKKDDDEDENPDQEKLLPKSPERQVLNVPVPSDSE	2677
QY	1803	ITTEITIKRRDVGPIGIREYCIKILICPIGVPETKPMP---TPORKGLRSSALRPKR	1859
Db	2678	ISEIE-LEYSPTSEKSESQCYTETVRTTTVTREYLDQPSVTRSRSPSEHDIISOYAP	2736
QY	1860	ETPKQTPGVIIETWVA-----EELELWEIRAFARVERKEKAQAVEQQAOKR	1906
Db	2737	ESPVEQDPVVEKTTTVIRQYHDEPPQEIEEQTIPEEVTVLREVYESPEGDEPEQH--Y	2793
QY	1907	LEQOKPTVATST-----TSPSSSTISLIS---PAQKVMVAIPISGVTTGT-----	1949
Db	2794	IETKTTTTITKEVHVPEEDVOISPVHSETSSEKOLPADEQLDEEVTESTTTATVTRER	2853
QY	1950	-----KWLTTTKVGSPTVTT---FOONKNEHQ-----	1973
Db	2854	YEEPEVRPPSGSEADDESHAPKYMETT---TTTTVTREYVESEDEHQOQSVQRDESP	2910
QY	1974	-----TFATWVK-----QGOSSNGVVQOQKVLGIIPSSSTGTSTQFTTSPQR	2016
Db	2911	APSEDSVKHVIETKTTTTVTTEERYEPEDSHSPVSEDDVHGFVKTTTTTTTTVTHEHFEPE	2970
QY	2017	TATVTRPNTSGGTTSSQVITGQIRPGMTVIRTPLO-----QSTLGKAIIRTPV	2069
Db	2971	-----DPPSDEHVVESRYAGSPV---PSEEDSSREIETTTTTVTVTRHF	3013
QY	2070	MVQPGAPQOVMQO--IIRQGPVSTAVSAPNTVSST-----PGKSLSATSTSN	2117
Db	3014	ELEDQOEHVVEQSVSAGSPVPSEKSERVIETTTTTNTVTREHFEHEDDIPTIVTSHD	3073

Db 2479 YQTQANTNKKIML-----PLTAAGKPLKTVATKAAQAQAVKORQLKSGHQVKPIQAKL 2533
 Qy 1981 ----QGOSNGVQVQVQ--KVLG-----IIPSSGT-----SQOTTSF--- 2013
 Db 2534 QPHQHQHQOQOQVQPIVGMQNLQQLLFQSSQTQQAQPIILPQAQPNIIISFTVG 2593
 Qy 2014 -----QPR-TATVTRPNTSGG-----TTSNS 2036
 Db 2594 DGSQQLQYISIPTAGEYKPPQPTATPTPLTAPGAGATYLTQDASGNLVLTTPS 2653
 Qy 2037 --QVITGPQIRPGMTVIRTPLOQSTL-----GKAIRPVVMVQPGAPQVM----- 2080
 Db 2654 GLQMLTAQSLAQAPQVIGTLIQPOTIQLGAGDN-----OPGSGNOOPLLGGTGG 2704
 Qy 2081 -----TOIRGQPV-----STAVSAPNTVSST-PG-----QKSLTSATSTS 2115
 Db 2705 GSSGLEFATTSPOVILATOPMYGLEIVQNTVMSQOFVSTAMPGLMSQNASFAITQ 2764
 Qy 2116 NIQSSASPPRPQGGQVKLMAQLTQTHGGNGQUTV-----VIQGGQ--TTGQLOLI 2169
 Db 2765 VFOASKIEP-----IVDLFAGYVVLNNTGDASSAGTFLNAASVLOOQOTQDDTTQI--- 2815
 Qy 2170 POGVTVPGPQQQLMAQAMPNGTVQREFLPTATTATTASTTTTTVTSTTAAGTGEQROSK 2229
 Db 2816 -----LQANFQ-FQSVPTSGASTSMDYTSPVMWTA-----K 2847
 Qy 2230 LSPQMVOHQDKTLPPAQS--SSVGPAKAQPO-----TAQPSARPQPTQPSQAQPEVQ 2281
 Db 2848 IPPVTQIKRTNAQAKAGISGVKVPQPPQVNVKVLPTSIVTQOSQVQVNSNLKSOVK 2907
 Qy 2282 TOPEVQQTIVSSVHPSEA-----OPTHAQSSKPOV-----AAQ 2315
 Db 2908 GKAAAGTGTTCGA--PPSIASKPQKKTNMIRPIHKLEVKPKVMKPTPKVQNONHSLQQQ 2966
 Qy 2316 SOPQSNVQOSPVV--QSP-----SOTRIREP-----PSQLSPGOSOVOT 2356
 Db 2967 QQQPQLQQLIPAVVNVQPKVTISOQIRIPAQTQQOQLOQAQMIHIFQQOQPLQOQOV 3026
 Qy 2357 TTSQPI---PIQPHTSLIQPSQGPQSQPQVQSSQTSLSSQTLNQVSVSPSR----- 2407
 Db 3027 QPSMPIITLAEAPVQVQFVMEPQALQEQELANRVQHFSTSSSSSSNCSLPTNVNPMQ 3086
 Qy 2408 -----POLQIQOPQ-----QVTAVQLOQOVQ--VLSQIQSQVVA 2441
 Db 3087 QQAPSTSSSTRPTNRVLPMQORQEPAPLSNECPVVSPTPKPVEQPIIHOHTSASVS 3146
 Qy 2442 QIQAQQSGVPOQIKLQPLQIQOSSAVQTHQIQNVTVQAAASVOEL----- 2488
 Db 3147 KQYAKQSTLSPV--YEAELKVS SVLES--IVPDVTMDALIEQPVQTSIYTBGLYEKN 3201
 Qy 2489 ----QRVQOLRDOQKQKQ--QJIEINVNTPSKLLIRVEIQOVVMKHNABIEH----- 2537
 Db 3202 SPGESKEQLLLQOQREQLNQVLNNGYLLDKHTFQVPEPMDTVYREEDLEEEDEDD 3261
 Qy 2538 --LKQKKS-WTPABRENQRMVNCQVWKYILDKIEE----- 2573
 Db 3262 FSLKMTSACNDHMSDESEPAVVKDKISK-ILDLNTDCCADSTATATTMEVDASAGYQ 3320
 Qy 2574 -----KQAAKKKREESVEQKSKQNAKLSALLFKHKEQLRAELILKRALLD-K 2622
 Db 3321 MVEDVLATTAQASAPTEFEQALETAAVEAAATYINEMADAH-----VLDLK 3367
 Qy 2623 DLQIEQBELKRLDKIKKEDLMOLAQATAVAAPCPVPTVLPAPPAPPPPPPPG 2679
 Db 3368 QLQNGVELELRR--KEEQTVSQEQQSKAAIIVP--TAAAPPEPQIQEPKKTG 3419

RESULT 8

T34418

hypothetical protein F12F3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34418

R:Fulton, B.; Wohldmann, P.
 submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of C. elegans cosmid F12F3.
 A:Reference number: 221521
 A:Accession: T34418
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-3488 <FUL>
 A:Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
 A:Experimental source: strain Bristol N2; clone F12F3
 C:Genetics:
 A:Gene: CESP:F12F3.3
 A:Map position: 5
 A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 3.1%; Score 458.5; DB 2; Length 3488;
 Best Local Similarity 18.0%; Pred. No. 4.1e-09;
 Matches 595; Conservative 494; Mismatches 1226; Indels 983; Gaps 138;

Qy 84 DIPLEFPKSS-EDLAMPNEHIMVI-----AIYEVLRNFGTVLRSLSPRFEDF 131
 Db 5 DFPVANEDKMAFEGFTKGSEHSFNVTINDIRKDLGGKLVFEAKNDYGV----- 52
 Qy 132 CAALVSQEOCTLMAEMH-----VLLKAVLREEDTSNTTFPADLKDSVNSTLY-FIDGM 185
 Db 53 -----DKCTILLDVRDGSFIEDYSEIHSAEIQNSV-GDVQVQKEGTAKLGRVDGF 104
 Qy 186 TPWEVLRYCESDREYHHVLP---YQEAEDYPYGPVENKIKVLOFLVDQFLTNIAREEL 242
 Db 105 PLPEL--IWKNGAIEDMMVPSTKYQ-LDYHSDGEFEARIANCTFEDD----- 149
 Qy 243 MSEGVIQYDDHRCVCHKLGLCCGTCSAVYHLECVKPPLEE-----VPDE 289
 Db 150 -----DDYSLIVENLAGV---DSCNFQVFVDCNEYPDDDEHFNRRRLQGRVMEAS 198
 Qy 290 WQCEVCVAHK-----VP-----GVTDCAVEI 310
 Db 199 SDSELDDAKKRRRIKRVVERPNAPRLTQLIPRFDKILSDHDAIEGENVMMVETL 258
 Qy 311 QKNPYIR-----HEPIGYDRSRKRYWFLNRLIITEEDPENENEK-----K 351
 Db 259 GEPEPQVRFYRDGKLIDGSGDRMEVHEDEMRKHWLLKIDCKDEAEYACQAINVAGE 318
 Qy 352 IWYYS-TKVQJAEILCDLDKYWEAELCKILEEMREEIHRHMIDETJTNKARSGNKSFL 410
 Db 319 AWCFSDVVVMHSESRDDKSDVEVDDSTVLEKKDD---GDDKSKPKTKKKIKKKETP 375
 Qy 411 AAANEELIESTRAKKGDIDNVKSPETE---KDNKETENDS-----KDAENREEFEDQ 461
 Db 376 ESEQVTAEPPEQQKISEVD-VQSVATEVGVAKKXPDAEKPTDLSKAKKDSKSKSDEPEA 434
 Qy 462 SLEKDSDDKTDDDPQEQKSEPTVEVGDKGNSVANLGDNTNATSETPSESGSPVGC 521
 Db 435 STEEKSTETKNTDKTSKSAKKTVKPK-----KEVTGRPLEAKKPVEDKKDASQ 485
 Qy 522 LSETPDSNMAEKKVASLQPD--VPEPNKTCSSNTSATTTSIQPNLNSNSSE--- 576
 Db 486 PSSKSSPPPTDGKKKQIPKALFIPDEISSRFGDPSTMHSETNITTTIRREGSADAKT 545
 Qy 577 -----LNSQS-----ESAK-----AADDPENGERSH--TPVSIQEEIVGDFTS 614
 Db 546 PLVEPLSASVMKYFTTLVESAKEAEFSFKRSRSETPDDKSKKEGLPPAKKSEKDEYTA 605
 Qy 615 EK-STGELSESPGAGKASGSTRITLRLNPDLSQLSKSQVAAAHAANKLFKEGKEV 673
 Db 606 EKQSTEALIES-----KKKEVDESKISEQQSDKKNSSEVVGVPEKAAGETKKDSEIEEV 661
 Qy 674 -----LVVNSQGEISRLSTKKKEVI-----MKGNNINNYFKLQGEQKYRVYHNOYSTNS 720
 Db 662 PKKTKIKKTEKSDSSISQKSNVLKPADDDKSKDDVDYDKSKKTTEDOTKVATDSKLEKA 721
 Qy 721 FALNKHQHRE-----DHDKRRHLAHKFCULTPAGEFKWNGSVHGSKVLTITSLRLTITQLEN 776

Db 722 ADTTKQIETETVVDKSKKVLKKK--TEKDSF-----ISQSETPPVPEP 766
Qy 777 NIPSSFFHPNASHRANWIKAVQMSKPREFALAL-AILECAVKPVV--MLPTWREFLGH 833
Db 767 TKP-----ASEAQKIAEVNKAQKQEVDDNLAREAEVAAKKADEKLEAE- 814
Qy 834 TRLHMTSIEEBEKEKVKKEKOEETMQQTWVYKTPPVKHQVWKQGEERYVTGYG 893
Db 815 ANIKKTAEEVAAKQKQEKDQLKLETEVSSKSAEKLKLEKQAOI--KKAAB- 865
Qy 894 GWSWISKTHYRVPKPLPGNTVNVYRKSLEGTKNNMDNM----- 933
Db 866 -----ADAVKKQKELNEKNEKLEAAKSAADKLKEESAASKKVESEVPK 912
Qy 934 DESDKRCSRPKIKITEPSEKD-EVKGSDAAKADQ--NEMDISKITEKKDQDKELL 990
Db 913 GEEKKTAGEKTVOVESEPTSKKTIOTKDVGATEPADETPKKKIIKKKTEKSDSSISQKS 972
Qy 991 DSDSDKPKC---EEP-----MEVDDDMKTESHV-----NCQBS 1020
Db 973 ATDSEKYSKQEQDEPTKPAVSETQMVTEADKSKQKETDEKLDLDAEIAAKTKQEADEK 1032
Qy 1021 SOVD-----VNVSE-----GPHLRTSYKKKTKSSKLDGL-LERRIK-----OF 1058
Db 1033 SKLDAQEKIKVSEDDAARKEKELNDKLESEATKKAADKLEBQAKAAAEVAA 1092
Qy 1059 TLEEKQREKIKLEGGIKGIGKTSNKKNLSESPVITK---AKEQCQSDSMRQEQSPNA 1115
Db 1093 AKKQEKDEQKLD-----TEAASKKAAAEKLEKQAKKAAAGDAVKKQKELDE 1144
Qy 1116 NNDQPEDLIQCCSDSSVLMSDPSTHNNKLPKDRVLDVDSIRSPETKCPKQNSIEND 1175
Db 1145 KNK-----LEANKKSAAGLKIEESAASKQVTEQAKLDAQTKAKTAB--KOTKLEKD 1197
Qy 1176 IEKVSDIOLASQGPTEKSKTKGNDFIDDSKLASADDIGTLICKNKKPLIQEESDRTVS- 1234
Db 1198 -EKSTKESKETEYDEPKKK-----VLKKK-----TEKSDSSISQ 1232
Qy 1235 ---SSKSALHSSVPKSTNDR---DATPLSRAMPEFKLGD-----SENSTL 1276
Db 1233 KSETSKTVESAGSESETOKVADAARKQETDEKQKLEAEITAKKSADEKSKLEAESKL 1292
Qy 1277 ENSSDTVIODSSSEDMIVONSIESIQFRTQEDVEVLEPLKCELVSGESTGNCDRL 1336
Db 1293 KKAEEVAAKQKQEKDQLKLDTEAASKKAAAEKLEKQSHIK----- 1336
Qy 1337 PVKGTANGKPKSQOKKLEER-----PVNKCSDQIKLNTTKNN-----NENRESE 1383
Db 1337 --KAAEVDAVK--ROKELEEKQRLSEEAATKKADEKLEKQKKAEEIALIEIOKEQ 1392
Qy 1384 KKGQRTSTFQINGKDNKPKIYLKGECLKEISESRVVGNNVEPKVNNINKIIPENDIKSL 1442
Db 1393 KLAQEQSRLEDEAKKSAEKQKLESETKSKOTEAPKESVDKPKKVKLKKKTEKSD--SSI 1451
Qy 1443 TVKESAIRPFIN-GDVIMEDFN-----ERNSSETKSHLSSSDAEGNTRDSLETL 1491
Db 1452 SOKSKASAKTVDAAEATLESDFNLVEKKTQVKEQSPDETSATIKRPAQ-----KTEEI 1506
Qy 1492 PSTKESDSTQTTTSPASCPESNSVQVEDMEIETSE--VKKVT--SSPITSE-----EEN 1543
Db 1507 SKQDDGDEKKTIT-----DGKPKPKPDESEATPKKRVVKKTKQKSDSVASDASLADYSK 1559
Qy 1544 LSNDFIDENGLPINKNENVNESKRKTVITEVTMTSTVATESKTVIKVEKGDK----- 1597
Db 1560 LSDDVEEPKPKVLKKKT-----EKSDSVISE-TSSVDTIKPESE-VEIPTKAEQMIHLNR 1613
Qy 1598 ----QTVVSSSTENCAK-STVTTTTTKLSTPSTGGSDVILISVKEQSKTVVTTT----- 1647
Db 1614 FSTDSAVESEPKNAHKDTEKTTDDMMTRKSSA-----IFSDDEQSISSKTSGEGRRR 1667
Qy 1648 -----VTDLSLTITGGT-----LVTSMTVSKEYSTRDK-----VKL 1677
Db 1668 RRTGFAKFSADTLALRGDNVETEAEALLAEDDTVTWVNGKADADLNSRCHESHMTFFRT 1727

1678 MFSRPRKTRSGTALPSYRKFEVTKSTKKSIFVLNDDLLKKLARKGGIRE-----VPYFN 1731
Db 1728 LIIDEVEPTDSMEITATCGTESHTTILKVEELPVDFVKYLPKRTSGKEGEVITSITLN 1787
Qy 1732 YNAKPAIDIWPPSPRPFGITWRYRLQTVKSLAGVSLMLRLNLWASLWDDMAA-KYPPG 1790
Db 1788 HPIDISKVW-LKDGKP-LEINKOYSIDTVG--CSVSLTLR-----RAKIEDSGKYKVVCD 1839
Qy 1791 GSGTIT-----ETSETEITTT-----EIIKRRDVGVP-----GIRFEC 1824
Db 1840 GUDCSTHLSIOQKPVLKNNVSETKPVITVDKDDQSLVAYDSNPEASFMTVDGKLEFD 1899
Qy 1825 IRKIICPI-----GVPETPK---ETPTPQRKGLSSALRPRPTPKQTGPVIE- 1871
Db 1900 GRSRDVDDGLKTRGVSKTAGEYEYKLNKNEFEVAQKFDVKVNDTPSAGDVSVVK 1959
Qy 1872 -----TWAEELWELEWRAFAERVEKEKA-----OAVEQQAQKR-----LEOK 1911
Db 1960 AESDCLHIEWTAPTEDNGAEVTSYV--LEKESGRKKFKHKVATVNGKKTSTVDDLEIET 2017
Qy 1912 PTVIATSTTSP-----TSSTTSTISPAQ--KVMVAPISGVT- 1947
Db 2018 PIVIRIAAVNKFGTGEFETKPVQSGSPQVPTVEFPPTIDNVTSTSCSLSWPKPIEDGG 2077
Qy 1948 -----GTMVLTITKVGSPATVTFQONKNEFHOTFATVVKOG-OSN 1985
Db 2078 SPVYGYDVYKRENEGEWQKMGEEELVFTESFNVA---LSSGKEYEFKIEACNEAGLRN 2134
Qy 1986 SGVVOVQOKVLGIIPS-----STGTSQOFTSFQPRPTATVIRPNTSG----- 2028
Db 2135 SNVSKKLTVGLVPEIILLDMPMVKVLDNDKVEVTKSDEGEFFVQVYKSDGSSIASVD 2194
Qy 2029 SGGTTSNSO-----VITG----- 2047
Db 2195 IGGPSESAAATSKCIIDGLREGIPYVFRVAARNOHGTGEFSEPTIPVVVLADDAAPRLKA 2254
Qy 2048 MTVIRTPLOOSTL-----CKAIIRT---PVMVQGAPOQVMTQIIRG 2086
Db 2255 IKPVKIPKGELELRLECHAAGHAPAEYIWKDKEIIPDENTEIVNEGSMALIIHELAG 2314
Qy 2087 QPVST-AVSAPNTVSTPGQKSLTSATSTNSIQSSASQPPRPOQG-QVKLT----- 2135
Db 2315 EDVGLYKVLVENTHGTAESEAEVIGISDVRAHFNSSSELTEIEEGHDIETCEVSEDAV 2374
Qy 2136 -----MAQLTOL-----TOGHGNGOGLTV--- 2155
Db 2375 VNMVKGKLVASDRVQFYAMARKRTLRIKGTSDADSGVYKCTETDGRSRTGEVIVNEQ 2434
Qy 2156 -----IQOGQOT-----TGOLILPQGVTVLPQGGQOLMOAAMPNGTV 2193
Db 2435 EPHILVGPDAIVKQGETMVLFCETSKPVKVKFKNGVEIWPQMNKAIMENDCKRATL 2494
Qy 2194 QRLFTPLATTATTATSTTTTSTVTAAGTGEORQSKLSPO-----MQVHQDKTLPPAOS 2248
Db 2495 EIKNFKDHIDIGAYTASVSEKETSAPAKLVFVAPNLIIPTDIRDGTVHAGNEFDFA--- 2551
Qy 2249 SVGPAKAQQTAPSAARPOQPOPOPAOPEVOTQPTVTSVSHVSEAPQTHAOS 2308
Db 2552 -----VEFSGFPIPTHLNNGTPLKAIIVVTEYDSDSVSRM---KDVTLDSNG 2597
Qy 2309 KPOVAAQSQPSNVQGSQSPRVQSPQTRIRSTPSQLSPGQOQSVQTTTSTOPIQPH 2368
Db 2598 TVRVIAES-PLGQCIKEIPLKIID-----KPSAPCDL-----QFREVTEDSVFLSWOP-- 2644
Qy 2369 SLOIPSQGQPOQPOQVQSSQTILSSQTLNOVSVSSPS-RPOLQIOQOP----- 2417
Db 2645 --PLETNGAP-----LTGYIERKAVDNNRWRPCGVKPYKLTFAEDLFCN 2689
Qy 2418 -----QVIAPVQLQOO-----VOVLSIQSVVAAIQAOQSGVPQOIKQLPIQOOSA 2467
Db 2690 QVYGFRLIAVNEGESEPCDITVDVLTLESSEPVSS--ESSELFVPKIALIRTP---QVTV 2744

cell proliferation antigen Ki-67, long form - human

C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins.

A:Reference number: A48666; MUID:94043435; PMID:8227122

A:Accession: A48666

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-3256 <SCH>

A:Cross-references: EMBL:X65550; NID:g415818; PIDN:CAA46519.1; PID:g415819

C:Superfamily: kinase interaction domain homology

C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat

F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.9%; Score 439.5; DB 2; Length 3256;

Best Local Similarity 18.3%; Pred. No. 1.9e-08;

Matches 579; Conservative 432; Mismatches 1103; Indels 1049; Gaps 148;

QY 300 VPGVTDCAEIKKPKYIRHEPIGYDRSRKRYWFLNRLIIEEDPENENKKIWT----- 354

DB 221 VP-TTQCLDNKKNE-----SPFWKLYESVKKELDVKSQENVLQYCRKSG 265

QY 355 -----YSTKVQLAELIDCLDKDYWEAELCKILIEEMREEIHRHMDIPEDLTNKARGSNKSF 410

DB 266 LQTDYATKESADGLQ-----GETQLL-----VSRKSRPKSGSGHAVA 304

QY 411 AAAN-EEILESIRAKGDIDNVKSPET-----EKDKNET-----ENDSKDAENR 455

DB 305 EPASPEQELDNQNGRQVESQTPSKAVGASFPLYEPAKMTPVOYQQOONSPOKHKNK 364

QY 456 EEF-----EDQSLKLD-----SDKTPDDDPDEQCKSEETVEGDKNSV--ANLGDNTN 504

DB 365 DLYTTGRRESVNLGSEGFAGDKTLTPKLSRNRTPAKVEDAADATKPNLSKTRG 424

QY 505 A-TSEETSPSE---GRSPVGCSET-----PDSSNNAEK-----KVASELPQDVPE 546

DB 425 SIPTDVEVLPTETEIHNPPFLTLWLTVQVERKIQKDSLSKPEKLGITAGOMCSGLPLSSV 484

QY 547 ENKTCESNTS-----ATTSTIOPNL--ENSNSSSLNSSQSESASAA----- 588

DB 485 DINFGDSINESEGLPLKRRRVFSGGHLRPELFDENLPNTPDKRGEAPTAKKSLVMHTP 544

QY 589 -----DDPE-NGERES-----HTPVSITQEEIVG-----DFTSEKSTGELSESPG 626

DB 545 PVLKIKIIEQOPQPSQKQESGEIHEVVKQAQSLVISPAPSPRKTVPASDQRRRSCKTAPA 604

QY 627 AGKAS-----GSTRITRLNPD-----KLSQLKSQVAAAHANKLFKKGKVL 674

DB 605 SSSKSQTEVPKRGGERVATCLQKRVISRSQSHDILQMICSKRRSGASEAN-----LI 656

QY 675 VVNSQGEISRLTKK--EVIMKG--NINNYFKLGQBECKYRV--YHNOYSTN-----S 720

DB 657 VAKSWADVVKLGAKOTQTKVHKGPQSRNMKQRAPATPKPKPVGEVHSOFSTGHANSPECT 716

QY 721 FALNKHQHRREDHDKRR--HLAHLFCGLTPAGEFK-----WNGSVHIGSKVLITLIRL 769

DB 717 IIGRAHTEKTVHVPARYVLANFISNQMKDEKEDLSGIAEMFKTPVKEQPOLT-STCHI 775

QY 770 TITQLENNTPSSF-----FHPNASHRANWTKAVOMCSK--PREFALALALECAVK 819

DB 776 AISNSENLIGKQFGQDTSGEELPPTSSEFGNFFVFAQNAAKQPSD-----KCSAS 827

QY 820 PVVMLPIWREFLIGHT-----RLHRTSIERE-----EKEK 849

DB 828 P-----PLRQCIRENGNNAKTPRNTYKMTSLETKTSDETETPSKTVSVNRSGRTERFN 883

QY 850 VKK--KEKQEEETM-----QOATWYKTYTPVKHQVQKQGEYRVYTGGSW 898

DB 884 IQKLVPESSEETNTEIVEICILKRGOKATLLQ-----RRGENKEI----- 925

QY 899 SKTHYRVFVPKLPGNTVNVYRKSLGKTKNNMDENMDESDK-----RKCS--RSPKKIKIE 951

DB 926 -----BRPET---YKENTELKEN--DEKMKAMKRSRTWGQCAPMSDLTKLSL 970

QY 952 PDSE--KDEVKG-----SDAAKADQNEMDISKITEKKQDVQKVELLSDSDSKPCKE-EP 1002

DB 971 PDTLMKDTARGQNLLOQDHAQAPKSEKGIITKM-----PCQSQUP 1012

QY 1003 MEVDDDMKTESHVNCQESSQVDVNVNSE-----GFHLRITS-----Y 1038

DB 1013 ----EPINTPHTTKQQLKASLGKVGKVELLAVGKFTRTSGEHTTHREPADGKSI RTF 1068

QY 1039 KKKTK-----SSKLDGLLERIKQFTLEEKORLEKIKLEGKIGKIGKIGKIGKIGKIGKIGK 1093

DB 1069 KESPKQLDPAARVTGM--KKWPRTPKEAQSLEDL---AGFKELFOTPGPSEESMTDEK 1123

QY 1094 VITRAKEGCSDSMRQEOPNANDQPE-DLIOGCSQSDSVLRMSDPSTHTNKLYPKDR 1152

DB 1124 T---TKIACKSPPPESVDPTPTSTQWPKRSRKADVEEFELALRLKLPKSAGKAMLT PKPA 1180

QY 1153 VLDDVSI-----RSPETKCPKONSIE-----DIEBKVSDL 1183

DB 1181 GGDEKDIKAFMGTPVQKLDLAGTLPGSKRQLQTPKEKAQALDELLAGFKELFOTPGHTEEL 1240

QY 1184 ASRQ-----EPTKSKTKGNDFDDSKLASADDIGILI--CNKKPLIQE 1227

DB 1241 VAAGKTTKIPCDSPQSDPVDPTSTKQR-----PKRSIRKADVEGELLACRLNMP----- 1290

QY 1228 ESDTIVSSSKSALHSSVPKSTNDRD-----ATPLSRAMDFEGKLGCDSESNSTLSENSDT 1282

DB 1291 -----SAGAMHTPKESVGEKDIIFVGTVPQKL-----DLTENLTGS 1329

QY 1283 VSIQDSEEDMIVQNSNIESISEQFTRQDVVEVLEPLKCELVSGESTG--NCDRLPVLKGT 1341

DB 1330 KRRPOTPREAQALEDLFGFKELFOTPGHTEEA-----VAAAGTKTKMPCESSPPESAD 1382

QY 1342 EANGKKPSQOKKLEBPVYNKCSQDLKANTTDKNNENRESEKKQORTSTQINGKDNKP 1401

DB 1383 TPTSTRROPKTPLEKRDVQKELSALK-----KLQTSGETTHTDKVPGEDKS 1430

QY 1402 KIYLGECLEISESRVVSNGV--EPKVNINIKIIPENDIKSLTVKESAIRPFINGDVIME 1460

DB 1431 INAFRETAQKLDLPAASVTGSKRHPKTK--EKAQPLEDLAGW--KELFOTP-----VCT 1480

QY 1461 DFENRNSSETKSHLLSSSDAEGNYRDSLETLPSTKESDSTQTTPTPSACSPSNSVNOVED 1520

DB 1481 DKPTTHTKTTIACRSQPD-----PVDPTSSKPSQSKRSLRKVD 1519

QY 1521 MEIETSEVKVYTS-----PITSSEENLNDFDENGLPINK--NENVNGESKR 1568

DB 1520 VEEEFALRKRTPSAGKAMHTPKPAVSKEKNIYA--FM---GTPVKLDLTENLTGSKRR 1574

QY 1569 KTVITEVTMTSTVA-----TESKTVIKVERGDKQTVVSVSSTENCAKSTVTVT 1616

DB 1575 LQTPKEKAQALDELLAGFKELFOTRGHTEESMT--NDKTAKVACKSQPDLDKNPASSKR 1631

QY 1617 TTVTKLSTPSPGGSDVLIISVKEQSKTVVTTVTDSLTGTTGGTLVTSMTVSKYSTRDKVK 1676

DB 1632 RLKTSLG--KVGKVELLAVGLKTQTSGETTHTTPTGDCGSKMAFAPESKQIILDSAS 1689

QY 1677 LMKFSRPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLPNDDL----- 1715

DB 1690 LTGSKRQLRTPKGSKEVPEDELAGFTLFTQPSHTKESMTNEKTKTVSTRASQPDLDVPT 1749

QY 1716 -----KILARKGGIREVPYFYNNAKPALDIPWPSRPRTFGITWRYLQTVKSLAGVSL 1769

DB 1750 SSKPQPKSLRKADTEE-----EFLAFKQTPSAG-----KAMHTPKPAVGEEK 1793

QY 1770 MLR--LLWASLRWDDMAAKVPPGGSTRTTSTETTTTTEILKRRDVGPYGR----- 1820

DB 1794 DINTFLGTVPQKLDQGNLP--GSNRLQTRKEKAQALELT-----GFRELFTQPTCT 1844

Qy 1821 -----FEYCIRKIIC--PIGVP-ETPKETPTPQKGLRSS-----ALRPKRP----- 1859
Db 1845 DNPTADEKTKIILKSPQSDPADPTNTKQRPKSLKADVEEFLAFRLKLTTSAGKAM 1904
Qy 1860 ETPKQTVIETWVAEELELEWIRAF-----AERVE-----KEKAQAVE 1900
Db 1905 HTPKAA-----VGEK-----DINTFVGPVEKLLDGLNLPQSKRRPQTPKEKAKALE 1952
Qy 1901 QOAKRLEQOQPTVIATSTPSSTSTISPAQKVMVAPISGSSVTGTRKVLTKTKYGP 1960
Db 1953 DLAGEFKELFTPGHTEESMTDDKITEVCSKSPQDPVKTP-----TSSKORLKISLGK- 2005
Qy 1961 ATVTFOQKNFHTFAWTKVQGSNSGVVOQKVL--GIIPSTGTSQOFTFSFQPRTA 2018
Db 2006 -----VGKVEVLPVGLTKTQSGTKTQT----- 2028
Qy 2019 TVTIRPNTSGSGT-----TSNSQVI-----TGQIRPGMTVIRTP-----LQOSTLG 2061
Db 2029 -----HRETAGDGKSIKAFKESAKQMLDPANKYGTGMRWP-----RTPEEAQSLDLAGF 2079
Qy 2062 KAIRTPVMVQPAQOQVMTQIIRGQPVSTAVSAPNTVSTPGQKSLTSATSTSNIOSSA 2121
Db 2080 KELFTQPDHTEESTDDKTKI-----ACKSPPPESMDTPTST----- 2117
Qy 2122 SQPRPOOGQVKLT--NAQLTOLTQHGCGNGLTVVIOGOGTTCQLQIP-----QGVTV 2175
Db 2118 RRRKPTPLGRDRDIVEELSALKQLTQ-----THTDKVPGBDCKGINV 2159
Qy 2176 LPGGQOQLM-----QAAMPNGTVQ-----RELFTPLATTATTATTTTTV 2215
Db 2160 FRETAKQKLPAAVSATGSKQKQTPKGAQPLEDLAGLKELFQTPVCTDPTTHEKTKI 2219
Qy 2216 STTA-----AGTGB--QRQSKLS--PQMVHQDKTLPAPQSSSVGPAKAQPOTA----- 2260
Db 2220 ACRSQPDVGTPTIFRQSKSLRKADVEEESALRKRTPSVKGAMDTKPKAGDEKDM 2279
Qy 2261 -----QPSARPQQTQPSQAQPEVQTPQVQTTVSS-----HVPSEAQT 2303
Db 2280 KAFMGTVQKLDLPNGNLPGSKRWPQTPKE-----KAQALEDLAGFKELFQTPGTDKPT 2332
Qy 2304 HAQSSKPAQAQSQPOSNVQGSFVRVQSPSQTIRPS-----TPS- 2344
Db 2333 TDEKTKIACKS-PQ-----PDPVDTASTKQRPKRNLRKADVEEFLALRKRTPSA 2383
Qy 2345 -----QLSPGQSQOVQTTTQIPIQHTSL-QIP-SQOGPQSQP----- 2382
Db 2384 GKAMDTKPAVSDKKNINTEFVTPVQKLDLGNLPQSKRQTPQPKAEALDLVGFEL 2443
Qy 2383 -QVQSSQTLLSGQTLNOVSVSSPSRPLQIQPOQVIAVP-----QLQOQVQLSOI-- 2435
Db 2444 FQTPGHTEESMTDDKITEVCSKSPQESFSTSRSSKORLKIPLVKVDKKEEPLAVSKLTR 2503
Qy 2436 QSQVVAQIQAGQGVPOQIKL--OLPTIQOQSSAVQTHQIONVVTVQAASVOEQORVQO 2493
Db 2504 TSGETTTHTEPTGDSKIAFKESPQILDPA-----SVTGSRRQLRTRKEKARA 2555
Qy 2494 LRQOQOKKK-----QOQIEINVNT-----PSKLLIKVELIQ-----KQVMKH 2531
Db 2556 LEDLVDFKELFSAPGHTEESMTIDKNTKIPCKSPPELTDATSTKRCPTRPKEVKEE 2615
Qy 2532 NAVIEHLKQKSMT-----PAERENQRMVNCVMKYIILDKIDREKQAANKRKEES 2585
Db 2616 LSAVERLTQSGSTHFKEPASGDEGKIVL-----KORAKKKNPVE 2658
Qy 2586 VEQRKSKONA-----TKLSALLFKHKEOLRA-----EILKRA 2618
Db 2659 EEPSSRRPRAPKEKAQPLEDLAGFTELSETSGHTQESLTAGKATKIPCESPPLVDVTTA 2718
Qy 2619 LLDKDLQIEVQEEKLDKLIKKEKDLMLQAO-----ATAV 2653
Db 2719 STKRHLRTRVOK-----VQVKEEPSAVKFTQTSGETTDADKEPAGEDKIGALKESAKQT 2773
Qy 2654 AAFCPPTVPLPAPPAPPPSP-----PPPGVQHTGLLSTPTPLVASQKRRKEEKD 2705

Db 2774 PAPAASVTGSRRRPRAPRESAQATIEDLAGFKDPAAGHTEESMTDDKTKIPCKSPELED 2833
Qy 2706 SSSKSKKKKMISTTSKTKKTKLYCICKTPYDESKFYIGCDRCQNWYHRCVGLQSEA 2765
Db 2834 TATSSKRRPRTRAKQVEVKEE--LLAVCK-----LTQTSG 2866
Qy 2766 ELIDEEYVCQOQSTEDAMTVLPTLTKDYEGKRLVLSLQAHKMAWPFLEPVDPN-DAPD 2824
Db 2867 E-----THTDKPEVGEKGTKA-----FKQAKRNVDAED 2897
Qy 2825 YYGVIKPEM-----DLATMEERVQR-RYYEKLTFEVAD 2856
Db 2898 VIGSRROPRAPEKAQPLEDLASFQELSQTTPGHTTEELANGAAD 2940
RESULT 12
S37431
ankyrin 2, neuronal long splice form - human
N:Alternate names: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythro
N:Contains: ankyrin 2, short form
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence.Revision 06-Jan-1995 #text.Change 13-Aug-1999
C:Accession: S37431; A39643; B39643; A40334; A49462; S14533; S14569
R:Chan, W.
submitted to the EMBL Data Library, September 1993
A:Reference number: S37431
A:Accession: S37431
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3924 <CHA>
A:Cross-references: EMBL:Z26634; NID:9406287; PIDN:CAA81387.1; PID:9406288
R:Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.
J. Cell Biol. 114, 241-253, 1991
A:Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal
A:Reference number: A39643; MUID:91302466; PMID:1830053
A:Accession: A39643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2077 <OT1>
A:Cross-references: GB:56957
A:Accession: B39643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1443,3585-3924 <OTT>
A:Cross-references: EMBL:X56958
R:Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Wa
Genomics 10, 858-866, 1991
A:Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.
A:Reference number: A40334; MUID:92009921; PMID:1833308
A:Accession: A40334
A:Molecule type: DNA
A:Residues: 463-474, 'PE', 477-495 <TSE>
A:Cross-references: GB:M37123; NID:g178647; PIDN:AAA62828.1; PID:g178648
R:Chan, W.; Kordeli, E.; Bennett, V.
J. Cell Biol. 123, 1463-1473, 1993
A:Title: 440-KD ankyrinB: structure of the major developmentally regulated domain and
A:Reference number: A49462; MUID:94075409; PMID:8253844
A:Accession: A49462
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3924 <RES>
A:Cross-references: EMBL:Z26634; NID:9406287; PIDN:CAA81387.1; PID:9406288
C:Genetics:
A:Gene: GDB:ANK2
A:Cross-references: GDB:127607; OMIM:106410
A:Map position: 4q25-4q27
C:Superfamily: ankyrin; ankyrin repeat homology
C:Keywords: alternative splicing
F:2-3924/Product: ankyrin 2, long form #status predicted <MAT>
F:2-1443,3585-3924/Product: ankyrin 2, short form #status predicted <MA2>
F:63-95/Domain: ankyrin repeat homology <AN01>
F:96-128/Domain: ankyrin repeat homology <AN02>

F;129-161/Domain: ankyrin repeat homology <AN03>																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Qy	854 EK-----KQEBEETMQQATWVKYTFPVKHQVWKQGBEYRVTVGGWSWISKTHVYRFPV	908	
Db	1999 EKGPILTQREAAQKTENQTI-----KRGQRLPVTG-----	2027	
Qy	909 KLPNTVNVYRKSLGCTKNNMDENWDESDKPKCSRSKKIKIEPDSKDEKVGSDAAKGA	968	
Db	2028 --TAESKRGVRVSSIGVKK--EDAAGGKEKVLSHKIPVQSVPEESHRESEVPKEMA	2083	
Qy	969 D-QNEMDISKITEKK-DQDVKELLDSDSKPCKEPEMEVDDDM-KTESHVNCQSSQVDV	1025	
Db	2084 DEQGMDLQISFDRKSTDFSEVIKQLELNDNKYQFRLSETEKAQLHLD-----QVLT	2138	
Qy	1026 VNVSEGFHLRTSYKK-----KYKSKLDGLLE-----RRIKQFTLEE	1062	
Db	2139 SPFNTTFFL--DYMKDEFPLALSLQSGALDGSSESLKNEGVAAGSPCGSLMEGTFOISSE	2196	
Qy	1063 KORLEKIKLEGIGIKGIGTSTNSSKNSLSESPVITKAKEGCQSDSMROBQSPNANNDQ---	1119	
Db	2197 SYKHE-----GLAETPETSPELSFSFKSEEQGETKSTKTETTTTEIRSEKHP	2247	
Qy	1120 -PEDLIQGCSSDSSVLRMSDPSHITNKLYPKDRVLDVVSIRSPETKCPKQNSIEND---	1175	
Db	2248 TTKDITGGSEERGAIV--TEDSETSTESFOKEATLG-----SPKDTSPKR---QDDCTG	2296	
Qy	1176 -----IEEKVSDLASR--COEPTKSKTKGNDFFIDDSKLASADDICTLICK	1219	
Db	2297 SCVSALAKETPTGLITEEAACDEGQRTFGSSAHKTQT-----DSEAQES-----TATSD	2344	
Qy	1220 NKKPLIQEESDPIVSSSKSALHSSVPKSTNDRDATPLSRAMDFEGKLCGDCSESNTLNS	1279	
Db	2345 ETKALPLPEASVKYTDGTESKPGQVIRSPQGLELALPSR-----DSEVLSAVADD	2394	
Qy	1280 SDTVSIQDSSEEDMIVQNSNESISBQFRTREODVEVLEPKCELVSGESTGNCERL---	1336	
Db	2395 SLAVSHKDSLEASPVLEDNSS-----HKTPDSLEP-----SPLKESPCRDLSLESS	2439	
Qy	1337 PVKGTENGKKPS-----QKKLEERPVNKKCS	1364	
Db	2440 PYEPKMGAGIFFSPHPLPAAVAKTELLTEVASVRSELRLDDPGSAEDDSLEQTSLMESGG	2499	
Qy	1365 QIKLKNNTDKNNENRESEKKGQRTSTFQINGKNRPKIYKLGECLEKSEISRVVSGNVE	1424	
Db	2500 KSPSPDPDTSSEEVSVETPKTVDVST-----PKPAVTHECAEEDD-----SENGE	2545	
Qy	1425 PKVNNINKLIPENDIKSLTVKESAIRPFINGDIVMEDEFNERNSSSTKSHLLSDAEGNY	1484	
Db	2546 KK-----RFTPEEMFKMVTIKIMFDELEQEAQKRDYKKEPKQEBS-----SSDDPDADC	2597	
Qy	1485 RDSLETLPSTKESDSTQTTTPSACSPESNSVQVDEMEIETSEVKKVTs-----	1533	
Db	2598 SVDVDE--PKHTSGDESGVPVLVTSSEKVSSESEPELAQLKKGADSGLLPEPVIRV	2656	
Qy	1534 ---SPITSEESNLNDFI-----DENGLPKNNENNVGES-----KRKT	1570	
Db	2657 QPSPPLPMSDMSNPSEVQFVWVSQYTFKMNEDTQEPGKSEEEKDSESHLAEHRHA	2716	
Qy	1571 VITEVTMTSTVATESKTVIKVEKDKQTVVSTENCACSTVTITTTTTLKSTPSTGGS	1630	
Db	2717 VSTE-----AEDRSYDKLNRDQDKICDGHCCCEAMSPSSSSARPVSSGLQSTGDD	2767	
Qy	1631 VDIISVKEQS---KTVVTTVTDSLTGTLVTSMT-----VSKEYSTRDKVKLMMKF	1680	
Db	2768 VD-----EQPVYIKESALQGTHEKDTGEFELDVSRAESPQADCPSESESSSSLP	2822	
Qy	1681 SRPKKTRSG-TALPSYRKRFVTKSKSIFVLPNDDLLKLARKGGIREVYFVYNAKPA	1739	
Db	2823 SEGKELDEDIASSTQKTEVTKTDETFENLPKD-----	2856	
Qy	1740 IWPYSPRPRTFGITWYRLQTVKSLAGVSLMLRLIWLASLWDDMAAKVPPGGGSTRTERS	1799	
Db	2857 -----CPQDSSSIT---TQT-----DRFSMDVPVSDLAENDEIY	2887	

Qy	754	GSVHGSKV---ITJSTLKLTTQLENNIPSPFFHPNASHRANWIKAVQMCKSPREFALA	810
Db	711	MSDTGVSLSANSLSERLOQTN-SGDPE	739
Qy	811	LAILECAVPPVMLPIWREFLGHRLHRMSTSIBREEKEVKK---KEKKEBEETMQ--	864
Db	740	-----PITTEILGEKVLSTRNAKQOQSDRYASPTLRRSIKHENTVQTP	785
Qy	865	-----QATWVKYFPVK---HOVMK-QKGEERYVTGYGOWSWISKTHYVRFPKLP	911
Db	786	KNVHNITLDEKTPVSETEPLKTASSYKSLRRSRELHRT-----LVETMNEKTEAVLA	838
Qy	912	GNTVNVYRKSLEGT---KNMMDENMDESDK-RKCSRPKKIKIEPQSEKDEVKGSAAKG	967
Db	839	ENTTARH---LRGTFREQVDQOVDNENAPORCKESG---ELSEGEKTSARRSA---	889
Qy	968	ADQENMDISKITEKKDQDKELLSDSDKPKCKEPMVEVDDDMKTESHVNCQE---SSQVD	1024
Db	890	-----RKQKPTKDLGS-----QWVTQADYAEELLSQOQGT	921
Qy	1025	VVNWSEGHILRTSYKKTKSKLGDLLERRIK--OFTLEK-----ORLEKIKLEGGI	1075
Db	922	IQMLEESMHM-----QNTSISEDQIGTEKKVNIIVYATKEKHSPTKPGKKAQPLEGPAGL	976
Qy	1076	KGIGKTSTNSSKNLSE-----SPVIT-----KAKEGCOSDSMRQEOS---	1112
Db	977	KEHFETPNPKDPITEDRTRVLCKSPQVTTENITNTKPKQTSISGKKVD-MKEESSALTK	1035
Qy	1113	-----PNANNODPEDL-----IQGCSGDSVL---RMSDPSHTTNKLYPKDRVLDDVIR	1160
Db	1036	RTHMGESRHNPKILKECEDIKALKQSENMELTSTVNGSKRTLGKSKKKAQPLEDLT--	1093
Qy	1161	SPETKCPKQ---NSTENDIEEKVSLASRGQEPKTS-----KTKGNDFDIDSKLASAD	1211
Db	1094	-----CFQELFISPVPTNIIKKI-----PKSPTQPVRTPTASTKRLSKTGLSKVD	1139
Qy	1212	DIGTLICKNKKPLIOESDITVSSKS---ALHSSVP-KSTNDRDA---TPLSRAMDFEG	1264
Db	1140	-----VROEPSTLGRKTKSPRAGPTPAPVQVENDCTAYMETPKOLESIEN	1186
Qy	1265	KLGCDSSENS---TLENSSDTVSIODSSEDMIVONSNE-----SISEQFRTREQD	1312
Db	1187	LTLGRKQSTPKDITGFQDSFOIPDHANGLPIVVKTKMFPNSQPQESALTRKSRQSR	1246
Qy	1313	VEYLE-PLKCELVSGE---STGNCEDRLPVRKGTEA---NGKKPSQOKKLEERPYNKCSDOI	1366
Db	1247	ASISKIDVKEELLESEHLOLGEVDTFQVSTNKNVIRSSRPAKK-----	1291
Qy	1367	KLKNTDKNNENRESEKKGORTSTFQINGKONPKI---YLKGECLKEISESRVVSNGVE	1424
Db	1292	KLDSTAGMNSKRMKSCSKNDNTPCLEDLNGFQELFQMPGYANDSLTTGISTMLARSPLQ	1351
Qy	1425	PKVNNIN-KIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETKSHLLSSSDAEGN	1483
Db	1352	PVYQINKLSLPIILKMDVTEE-----ISG-----LWKQSLGRVHHTQEOEDN	1396
Qy	1484	YRDSLETLSTKE-----SDSTQTT---TPSASCPSNSNVQVEDMEI---ERSEVKKVT	1532
Db	1397	AIKAIMEIP--KETLQTAADGTRLRQQTP-----KEKVOPLDEHSVFQELFQTSRYC	1448
Qy	1533	SSPITSEESNLS-----NDFIDENGLPINKNENVNGE---	1565
Db	1449	SDPLIGNKOTRSLRSQPQGFVRTPTRSKRLAKTSVGNIAVREKISPLSPQCATGEVH	1508
Qy	1566	-----SKRKTIVETVMTSTVATESKTVIKVEKGDQO-----TVVSSSTE	1605
Db	1509	IPIGPEDDTENKGV---KESTPQTLDSASRTVSKRQQAHEERPQFSGLFHPQELFQTP	1566
Qy	1606	NCAKSVTTTTTVTKLSPSPGSGVDIILSVKEQSKTVVVTTVTDSTLTGGTLVTSMTV	1665
Db	1567	ASGKDPVTVDTETTKIALQSPQPGHIIINPASMKROS-----NMSL	1605
Qy	1666	SKEYSTYRDKVLMKFSRPRKKTSGTALPSYKRFVTKTKSIFVLPNDDLKKLARKGGIR	1725

```
Db 2604 --SRQLRTHKQWSTLLKLLGDSKEITQ---ISDHSEKLAHDTILKSTQOQKP---DS 2655
QY 2552 NQRMIVCQVMK-----YILDKID-----KEEQAAKKR--REESVEQKRS----- 2591
Db 2656 VKPLRTCRVLRASKEVPKEVLVDRDHATLQSKSNPLLSPKRKSARDGSIVRTRALRSL 2715
QY 2592 --KQATKLSALLFKHKQLRAELILKRALDKDLOI-----EVOBELK-----RDLKI 2638
Db 2716 AKQBATDEKVPPEKKRAASSKRYVSEPVKMKHLKIVSNKLESVEEQVSTMKTEEMEA 2775
QY 2639 KKEKDLMLQAATAVAACPVPTP-----VLPAPPAPPPPPPPGVQHTGL-LSTPTLP 2692
Db 2776 KREN-----PVPQNSRYRKKTNVQPRKFDASAEVGLKKEKTMK 2819
QY 2693 VASQRKRREKDDSSKSKKKMMIS--TTSRETKRDKLYCICKTPTPYDESK 2741
Db 2820 TASQETELQNPDDGAKKSTSRQVSGKRTCLRSRGTTEMPQCEAEKTSK 2870

RESULT 14
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34434
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GBI>
A:Cross-references: EMBL:U08046; PIDN:AACT08090.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: x
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 2.8%; Score 419.5; DB 2; Length 2232;
Best Local Similarity 21.8%; Pred No. 6.6e-08;
Matches 339; Conservative 172; Mismatches 633; Indels 413; Gaps 63;

QY 1071 LEGGKIGTGTSTNSSKNLSESPVITKAKEQCQSDSMRQE---SPNANNDQPEDLIQGC 1127
Db 731 LESSTSG-ATTSSGSAGTTMTSP-----SQSSVSGSGSGTSPAASSTSGEMTSQGS 781
QY 1128 SQDSSVLRMDSPTHTNKLKPKRVLDDVIRSPETKCPKQNSIENDIEKVDLSARG 1187
Db 782 TQTPGSSVSTAAILTSTQ-----QSVSTNSPGSTVTRPST-----VSGSTSSG 825
QY 1188 QEPKSKTKGNDFFIDDSKLASADDIGTLICKNKKPLQIESDITV-----SSSKSALHS 1242
Db 826 STVTVGSTEAS---TSGSVASSSPAPS-TSQNPMPSTSSGSMITQSPYPSQSTSPVES 881
QY 1243 SVPKSTNDRDATPLSRAMDFEGKGCDSSESNTLENS---SDTVSIQDSSSEDMIVQNSN 1299
Db 882 STTSPGSPGTTLTST-----PSPSQSTTIGTSGSTSPGISTSEE-WTSQGST 931
QY 1300 ESISEQFTRQDQVEVLEPLKCELVSGESTGNCEDRLPVKGTENGKKPKSQOKKLEERP 1359
Db 932 QTPGSTGTVAQPTVSD-----STSSGS---TVTVGSTEGS-SSPIPSTSQNTNPS 979
QY 1360 NKCSDOIKLKNTDKNNENRESEKKGQRTSTFQINGKDNPKIYLKCECLKELISESRV 1419
Db 980 TSSGSSMTQTPQSQSPVESSTSGATSS-----GSP-----GTTLTSTSPSPSP 1027
QY 1420 S-----GNVEPKVNNINKIPE--NDIKSLTVKESAIRFINGDVMEDFNERNSSET 1470
Db 1028 SSTIGSSQGSTSPVSVTISQSGTETPGSTGTVPKPSI-----VSGSASSGSTATMGSTEA 1083
QY 1471 KSHLLSSDAEGNRYRDSLETLPSTK---ESDSTQTTTSPASCPSPSNVQVEDMEIETSE 1527
```

```
Db 1084 SS---TSGGSSTSPNPSQSTSPSTSGATSSPGSSGTTLTSTISPSQSSTIGSSQGSTSP 1140
QY 1528 VKKVTSSPTTSEENLNSDNDFIDENGLPINKNENVG-----ESKRKTVTIT 1573
Db 1141 VVSTTSGMTSGQSTQIFG---STGSVTVQSTGSGSTSGEITSGOQTOTPRSSJST 1196
QY 1574 EVTMTST---VATES--KTIVKEBKDKQTVVSVSTENCACKSTVTTTTT-----TV 1619
Db 1197 SPAISTSTQSVSTNPGSTVTQ-----PSTVRGSTS--SGSTVTGTGEGSTSGSSSA 1249
QY 1620 TKLST-----PSTGSDVILSVKEQSKTVVTTVTYDLSJ-----TTGGT 1658
Db 1250 TSLSSSPVPSTSQSPNFSSTGSGSTPTPNPSQSTSPVSVTTTGEMTSHGSTOTPTSTIGST 1309
QY 1659 LVSTMTVSKYSTKDRKVKLMKFSRPKTRSGPALPSYKFKVTKSKSLFVLPNDLKL 1718
Db 1310 VTQPTVSGSNSGSTVTI---GSSEASTSGSF-----KTSFSSISPVPTSS--- 1354
QY 1719 ARKGIREVPYFNNAKPALDIWYPSPRPTF-GITWRYRLQTVKSLAGVSLMLRLMAS 1777
Db 1355 -----PIPTTFASSTSGSTISDVSSVSTSLA----- 1382
QY 1778 LRWDDMAAKVPPGGSGSTRETSETETIITEIKRRDVGPIGRFYCIRKIKICPIGVET 1837
Db 1383 ---PLSSLP---STVPSTQSFSTSEGSKASSP-----VPSQ 1417
QY 1838 PKETPTQKGLRSSLALRPK-----RPETPKQGTGPVLIETWVAEE 1878
Db 1418 TSSPTNTGTSTESSTLLSSTISGTSQHTMTSKASSGSTSPSTNSQTGSTVWG----- 1471
QY 1879 LELWEIRAFERKEKEKAQVEQQAQKRLKQOKPTVIATSTTSPSTSTSTISPAQKVMV 1938
Db 1472 -----SSSTGCVSTSSASSTQPMSTSGSSAGSTVASTASPAASSTAPSSTG--TMS 1523
QY 1939 APISGVTGTGKMWLTTKVGP---ATVTFQONKFNHOTFATWVKQGSNSGVVOQOKV 1995
Db 1524 STSGTGSTTISESSTTASASQTGSTVTMGSSST-----SGVSTSSASSTQPM 1573
QY 1996 LGIIPSSGTSTQOTSTSPQPTATVTRPNTSGSGTSTNSQVITGPQIRPQMTVIRPL 2055
Db 1574 STQSSAGSTVASTAGLVSTIV---PSTGTMGSTSG-----TVGSTIS 1618
QY 2056 QOSTLGKAIIRTPVMQVQAPQVMTQIIRGQVSTAVSAPNTVSTSPGQ-----KSL 2108
Db 1619 ESSTTASASSQTGSTVTMGSSS-----TSGVSTSSASSTQPMSTSGSSAGSTVASST 1672
QY 2109 TSATSTNIQSSASQPPRQGOQVKLTMAQLTQLTQGHGNOGLTVVTOGQGT----- 2162
Db 1673 TGLVSTSTVPSSTGTMGSTSGTGSTVSTISESS--TAASASSQTGSTVTMGSSSTSGVSTS 1730
QY 2163 --TGLOLILQ-----GVTVLPGPGQQLMQAAMP-----NGTVQRFLEPLATT 2204
Db 1731 SASSQPMSTSGSSAGSTVVSSTASPAASSTAPSTGTMSSTSGTGV-----STMSQS 1786
QY 2205 ATTASTTTTTTAAAGTGEORQSKLSPQMOMVQDK----- 2240
Db 1787 STAAATTSHTGTVTGLS-----SPTSSNQMTSTSGSSVSGSTVASTAGLVSTVPSSTG 1842
QY 2241 TLPPAQSSVG-----PAKAQPTAQ-----PSARPQQTQPSQPAQVEVQVQPEVQ 2287
Db 1843 TMGSTSGTGSTTISESSTTASASQTGSTVTMGSSSTSGVSTSSASSTQPMSTSGSS 1902
QY 2288 TOTTVSSH-----VPSEAPPHAOS-----SKPQVAQSQPSNVQGSPPVRV 2330
Db 1903 AGSTVASTAGLVSTSTVPSSTGTMGSTSGSTVSTGTTISESSTAASSTQGTG-----STVTI 1958
QY 2331 QSPQTRIRPSTPQSPQSQSQVQTTTSQIPHTSLQIPSGQSQSQSQVQSSST-Q 2389
Db 1959 GSTSGT--NPSSPSRL-----SQI-TITPSPSQSTESTQTSILPSSSPSPSTHVSSTSGT 2010
QY 2390 TLLSGQTLN-----QVSVSPSPRPOLOIQOQPOQVIAVPLOQOQVVLSS 2434
Db 1959 GSTSGT--NPSSPSRL-----SQI-TITPSPSQSTESTQTSILPSSSPSPSTHVSSTSGT 2010
QY 2390 TLLSGQTLN-----QVSVSPSPRPOLOIQOQPOQVIAVPLOQOQVVLSS 2434
```

Db	2011 TMSSGATTSGDKMFLSTGTTTSSRGSLATTAPKPSVTCLFMYDTQSKEIDQ 2067	Db	582 QL-----EHEPPTKLPYK--GH--VSREYQLDNEVRDELPEYKGVHSRE----- 624
RESULT 15		Qy	868 WVKYTFPVKHQWQKGBEYRVTYGGWSWISKTHVYFVFKPLPGTNTVNYRKSLGKTK- 926
D71623	erythrocyte membrane protein pFEMP3 PFB0095c - malaria parasite (Plasmodium falciparum)	Db	625 -----YQLDNEV-RDELPEY-----EKGVSR-----EYQLDNEGPST 656
C:Species: Plasmodium falciparum		Qy	927 -----NMNDENMDSRKRKSRSPKKIKIIPDSKDEKGVKSDAAGK--- 967
C:date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000		Db	657 LKEYDQTELAKGKDTITNKPHEVDYDQSELAKG-KDITNKPHEVDYDQTELAKGKEV 715
C:Accession: D71623		Qy	968 -----ADQEMDISK---ITTEKDDYVKELLSD--SDKPCKEPEWE-VDDDMKTE- 1012
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;		Db	716 TNKPHENLEENETDLAKGKEVTNKPHEVDYDQSELAKGKDTITNKPHEVDYDQTEL 775
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.		Qy	1013 -----SHVNCQESSQVDYVNVSEGFHLRTSYKKTKSKLQDGLER-----RIK 1056
A:title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.		Db	776 AKGKEVTNKPHEENETDLA-----KGKEVTNKAHENLEENETDLAKGK 823
A:Reference number: A71600; MUID:99021743; PMID:9804551		Qy	1057 QFTLEEKORLEKIKLEGGIKIGKTSNNS-KNLSE-----SPVITAKEGC----- 1102
A:Accession: D71623		Db	824 EYTNKAHENLEENETDLAK--GKEVTNKAHENLEENETDLAKGKEVTNKAHENLEBYN 881
A:Status: preliminary; nucleic acid sequence not shown; translation not shown		Qy	1103 QSDSMRQEQSPNANDQPE-----DLIOGCSQSDSVLRMSDPST-----TNKLYP 1149
A:Molecule type: DNA		Db	882 EYDLAKGKEVTNKAHENLEENETDLAKGKEVTNKAHENLEENETDLAKGKEVTNKAHE 941
A:Residues: 1-2441 <GAR>	Query Match 2.8%; Score 417; DB 2; Length 2441;	Qy	1150 KDRVLDDVSIRSPETKCPKQNSIENDIEEKV-----SDLASRGQPTKSKTGNDFDID 1204
A:Cross-references: GB:AE001371; GB:AE001362; NID:g3845092; PIDN:AAC71809.1; PID:g384509	Best Local Similarity 18.6%; Pred. No. 9.1e-08;	Db	942 N--LEEN--ETDLAKGKEVTNKAHENLEENETDLA-KGKEVT-NKARENLEENY 992
A:Experimental source: clone 3D7	Matches 525; Conservative 424; Mismatches 1053; Indels 824; Gaps 132;	Qy	1205 SKLASADDTGLTICKKPKLIQESDTIVSSSKSALHSSVPKSTNDROATPLSRAM--- 1260
C:Genetics:		Db	993 TDLAKGKEVTNKAHENLEENETDLAKGKEVTNKA--RENLEENETDLAKGKEVTN 1047
A:Gene: PFB0095c		Qy	1261 -----DPEGLGCGDSESNTLENSDTSVSIQDSSE-----EDMIVQNSNESIS 1303
		Db	1048 KARENLEENYEEK--DYMKNNELQNGSD-GLKENAELKNKELNKGSDGLKENA-ELKN 1102
		Qy	1304 EGFRTREQD--VEVLEPLKCELVSGESTG---NCE---DRLPVKTEANGKPKSOOKKLE 1355
		Db	1103 KELRNKSGDGLKENAELKNKELQNGSEGLKENAELKNKELQNGKSE--GLKENAELKNK 1160
		Qy	1356 ERPVNCSDQIK-----LKNNTDKKNE-----NRBS 1382
		Db	1161 ELR-NKSGEGLKENAELKNKELQNGSEGLKENAELKNKELQNGSEGLKENAELKNKEL 1219
		Qy	1383 EKKGO-----RTSTFQINGKDN-KPKIYLGECLEKISESRVSVGNVPEKVNINK 1432
		Db	1220 QNGSEGLKENAELKNKELNKGSEGLKENVYTNNDLKNNDIQNKDLSNKMKNKELLNK 1279
		Qy	1433 IIPENDIKS-----LTVKESAIRPFINGDVIMEDFNERNSETKSHLLSSDAEGNYR 1485
		Db	1280 DISNKMKNKELLNKDLSNEDMKKNKELLNKDLSIGNMEQONTGLKNTPSKGOON 1339
		Qy	1486 DSLETLPSTKESDSTQITTPSASCPSNSVNOVEDMEIETSEVKVTSSTPITSEESNLS 1545
		Db	1340 TGLKNTPNEROQNTGLKNTPS-----EQQNTGLKNTPSEGOQNTGLKNTPNEROQNTGLK 1395
		Qy	1546 N---DFIDENGLPINKNENVNGESKRKTIVITEVTMTSTVATESKTVIIVKVGDKQTVVS 1602
		Db	1396 NTPSEGOQNTGL--KNTPIEGOON-----TGLKNTPS-----EQQNTGLK 1434
		Qy	1603 STEN-----CAKSTVTTTTVTYKLTSTPSTGSGVDIISVKESKTVVTTVTDSLTGTTG 1656
		Db	1435 NTPNEROQNTGLKNAANKGOQNTGLKNTPSKG-----QONTGLKNTPNEROQNTG 1484
		Qy	1657 GTLVTSMTVSKEYSTRDKVLMKFSRPKKTSGTAL---PSVRKFTKTKKKSIFVLND 1713
		Db	1485 ---LKNTPNEROQNTGLK-----NTPSEGOQNTGLKNTPNERQ-----QNT 1522
		Qy	1714 DLKKLARKG---GIREVPYFNNAKPAIDWIPSPRPTFGITWRYLQTVKSLAGVSL 1769
		Db	1523 GLKNASKGOQNTGLKNAP----- 1541

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2003, 16:10:05 ; Search time 53.7816 Seconds
(without alignments)
8207.607 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MYSEEEEDGDAEETQDSE.....KLKGFKASRSHNNKLQSTAS 2781

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14333	100.0	2781	21 AAY57453	Human transcriptio
2	14170	98.9	2907	21 AAY57452	Human transcriptio
3	13428	93.7	2759	24 AAO16418	Human nucleic acid
4	2695.5	55.7	22 AAB94078	Human protein sequ	
5	1557.5	10.9	976	22 AAB68735	Drosophila melanog
6	1474	10.3	328	21 AAB54364	Human pancreatic c
7	1030	7.2	191	24 ABU70846	Human adipocyte Se
8	982.5	6.9	238	20 AAY04323	Human secreted pro
9	981	6.8	237	22 AAU19446	Human diagnostic a

10	789	5.5	149	23 AAU78845	SCAN/KRAB protein
11	781	5.4	246	22 AAU31944	Novel human secret
12	672.5	4.7	997	22 ABB62887	Drosophila melanog
13	572.5	4.0	2768	22 ABB68397	Drosophila melanog
14	560	3.9	5533	22 ABB65772	Drosophila melanog
15	560	3.9	5560	22 ABB71160	Drosophila melanog
16	549.5	3.8	6815	22 ABB68811	Drosophila melanog
17	534	3.7	3080	22 ABB64877	Drosophila melanog
18	529.5	3.7	3257	22 ABB67502	Drosophila melanog
19	508.5	3.5	2897	22 ABB58514	Drosophila melanog
20	441.5	3.1	4498	22 ABB58595	Drosophila melanog
21	437	3.0	3021	24 ABB75873	Human secretory po
22	434.5	3.0	2977	22 ABB69480	Drosophila melanog
23	432.5	3.0	1878	22 AAM40239	Human polypeptide
24	432.5	3.0	2137	23 ABB39618	Staphylococcus epi
25	432.5	3.0	3266	21 ABA2491	Human ORFX ORF255
26	430.5	3.0	3684	24 ABR47592	Breast cancer asso
27	430	3.0	1795	22 ABB69806	Drosophila melanog
28	430	3.0	2586	22 ABB68878	Drosophila melanog
29	429	3.0	3726	22 ABB63947	Drosophila melanog
30	428	3.0	1727	22 AAB95554	Human protein sequ
31	423	3.0	2285	22 ABB63057	Drosophila melanog
32	418	2.9	4019	22 ABE13839	Human lung tumour-
33	418	2.9	4025	23 ABB69736	Human polypeptide
34	416.5	2.9	2665	22 ABB48336	Human liver peptid
35	416.5	2.9	2665	22 ABB28314	Human peptide #965
36	416.5	2.9	2665	22 ABB33490	Peptide #996 encod
37	416.5	2.9	2665	22 ABB18950	Protein #949 encod
38	416.5	2.9	2665	22 AAM54270	Human brain expres
39	416.5	2.9	2665	22 AAM66665	Human bone marrow
40	416.5	2.9	2665	22 AAM14533	Peptide #967 encod
41	416.5	2.9	2665	22 AAM26950	Peptide #987 encod
42	416.5	2.9	2665	22 AAM02259	Peptide #941 encod
43	416.5	2.9	2665	23 ABB36319	Human peptide enco
44	416	2.9	10431	24 ABU54861	Human CAL25 amino
45	415	2.9	3236	24 ABB96155	Human Ki-67 chromo

ALIGNMENTS

RESULT 1
AAY57453
ID AAY57453 standard; Protein; 2781 AA.
XX
AC AAY57453;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human transcriptional regulatory factor SEQ ID NO:10.
XX
DE Human transcriptional regulatory factor; TCOA1; BLAST detection;
KW bromo-domain; cell proliferation; cancer.
XX
OS Homo sapiens.
XX
OS WO9957143-A1.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-JP02340.
XX
PR 30-APR-1998; 98JP-0137631.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Jones MH;
XX
DR WPI; 2000-052940/04.
DR N-PSDB; AA239033.
XX
XX Transcriptional regulatory factor containing a bromo domain and gene
PT TCOA1 encoding it

Db	541	QDVPEPNKTCSSNTSATTTI	QIOWLENSNSSSELNSQSP	SASAKAADDPENCRESHT	600
Qy	483	-----VGDFKSEKNGELSE	SPGAGKAGSGSTRIITRLR	NPDSKLSQKSOQVAAA	534
Db	601	PVSIQEEIVGFTSEKSTGELS	SPGAGKAGSGSTRIITRLR	NPDSKLSQKSOQVAAA	660
Qy	535	HEANKLFKEKEVLVYNQSGEI	SRLSTYKKEVINMGNINNYF	KLQEGKYRYVINYSTNS	594
Db	661	HEANKLFKEKEVLVYNQSGEI	SRLSTYKKEVINMGNINNYF	KLQEGKYRYVINYSTNS	720
Qy	595	FALNKHORHEDHDKRRHLAH	KFCLTPAGEFKWNGSVHGS	KVLITSLITLITOLENNIPS	654
Db	721	FALNKHORHEDHDKRRHLAH	KFCLTPAGEFKWNGSVHGS	KVLITSLITLITOLENNIPS	780
Qy	655	SFLHPNASHRANWIKAVQMC	SKRPFALALALECAVKPVML	PWIREFLGHTLRHWT	714
Db	781	SFFHPNASHRANWIKAVQMC	SKRPFALALALECAVKPVML	PWIREFLGHTLRHWT	840
Qy	715	SIEREEKYKKEKKEEETMOA	TWVYTPVKHQVKQKEEYRV	TGYGWSWISK	774
Db	841	SIEREEKYKKEKKEEETMOA	TWVYTPVKHQVKQKEEYRV	TGYGWSWISK	900
Qy	775	THYRVFVKLPNGTNVNYRKS	LEGTKNMNMENDES	DKRCSRSPKKIKIEPDSEKDEVK	834
Db	901	THYRVFVKLPNGTNVNYRKS	LEGTKNMNMENDES	DKRCSRSPKKIKIEPDSEKDEVK	960
Qy	835	GSDAAKAGQONEMDISKITE	KKODVKELLDSDSKPCKE	PEMVEDDDMKTESHVNCQES	894
Db	961	GSDAAKAGQONEMDISKITE	KKODVKELLDSDSKPCKE	PEMVEDDDMKTESHVNCQES	1020
Qy	895	SOQDVVNVSEGFHLRTSYK	KTSSKLDGLLERIKOFTLE	EKKORLEKIKLEGKIGIK	954
Db	1021	SOQDVVNVSEGFHLRTSYK	KTSSKLDGLLERIKOFTLE	EKKORLEKIKLEGKIGIK	1080
Qy	955	TSTNSSKNLSEPVITKAKE	QCSDSMRQEOPNANDQ	PDLLOGCSDSSVLRMSDP	1014
Db	1081	TSTNSSKNLSEPVITKAKE	QCSDSMRQEOPNANDQ	PDLLOGCSDSSVLRMSDP	1140
Qy	1015	SHTTNKLVPKORVLDVDS	IRSPETKPKONSIENDIE	EKVSDLASRGOEPTKSTKGNDF	1074
Db	1141	SHTTNKLVPKORVLDVDS	IRSPETKPKONSIENDIE	EKVSDLASRGOEPTKSTKGNDF	1200
Qy	1075	FIDDSKLASADDIGTICK	NKKPLIQEESDITVSSKS	ALHSVPKSTNDRDATPLSRAM	1134
Db	1201	FIDDSKLASADDIGTICK	NKKPLIQEESDITVSSKS	ALHSVPKSTNDRDATPLSRAM	1260
Qy	1135	DPEGKLGCDSESNSTLENS	SDTVSIQDSSEEDMIVONS	ESISEQFRTREODVEVLEPLK	1194
Db	1261	DPEGKLGCDSESNSTLENS	SDTVSIQDSSEEDMIVONS	ESISEQFRTREODVEVLEPLK	1320
Qy	1195	CELVSGESTGNCEDRLPV	KGTGTEANGKPSQOKKLE	ERPVPNKCSDQIKLKNTT	DKKNENR 1254
Db	1321	CELVSGESTGNCEDRLPV	KGTGTEANGKPSQOKKLE	ERPVPNKCSDQIKLKNTT	DKKNENR 1380
Qy	1255	ESEKKGQRTSTFQINGK	ONKPKIYKGBCLKEISE	SRVVSNGVPEKVNINKIIPENDIK	1314
Db	1381	ESEKKGQRTSTFQINGK	ONKPKIYKGBCLKEISE	SRVVSNGVPEKVNINKIIPENDIK	1440
Qy	1315	SLTVKESATRPINGDVIM	EDFERNSSSETKSHLLSS	SDAEGNYRDSLETLPSTKESDST	1374
Db	1441	SLTVKESATRPINGDVIM	EDFERNSSSETKSHLLSS	SDAEGNYRDSLETLPSTKESDST	1500
Qy	1375	QTTTPSASCPESNVQVEDME	IETSEYVKVYTSPTITSE	ESNLSNDFIDENGLPIKNKE	1434
Db	1501	QTTTPSASCPESNVQVEDME	IETSEYVKVYTSPTITSE	ESNLSNDFIDENGLPIKNKE	1560
Qy	1435	NYNGESKRTVITEVTTTST	VATESKVIKVEKGDQKOT	VYSTENCASKVTTTTTTVT	1494
Db	1561	NYNGESKRTVITEVTTTST	VATESKVIKVEKGDQKOT	VYSTENCASKVTTTTTTVT	1620
Qy	1495	KLSTPSTGSGVDIIISVKE	QSKTVVTTTVDLSLTTT	TGGTLVTSMTYSKEYSTRDKVKLMKF	1554
Db	1621	KLSTPSTGSGVDIIISVKE	QSKTVVTTTVDLSLTTT	TGGTLVTSMTYSKEYSTRDKVKLMKF	1680
Qy	1555	SPKKTSTRSTALPSYRK	FVTSTKKSIFVLPNDDL	KLARKGIREVYFVFNNAKPALDI	1614
Db	1681	SPKKTSTRSTALPSYRK	FVTSTKKSIFVLPNDDL	KLARKGIREVYFVFNNAKPALDI	1740
Qy	1615	WPYPSRPRTFGITWRYRL	QTVKSLAGVSLMLRL	LWASLWDDMAAKVPPGGSGSTRTSE	1674
Db	1741	WPYPSRPRTFGITWRYRL	QTVKSLAGVSLMLRL	LWASLWDDMAAKVPPGGSGSTRTSE	1800
Qy	1675	TEITTTTEIIKRDDVGPY	IRFEYCIKIIICPIGV	PETPKETTPORKGLRSALARKPE	1734
Db	1801	TEITTTTEIIKRDDVGPY	IRFEYCIKIIICPIGV	PETPKETTPORKGLRSALARKPE	1860
Qy	1735	TPKOTGPVITETWVAE	ELELEWEIRAFAE	VEKEKAQAVEQOAKRLEQOKPTVIATST	1794
Db	1861	TPKOTGPVITETWVAE	ELELEWEIRAFAE	VEKEKAQAVEQOAKRLEQOKPTVIATST	1920
Qy	1795	SPTSSTTSTISPAQK	VWVAPISGSVTTG	TKMWLTTKVGS	PATVTFQONKNFHQTATWVK 1854
Db	1921	SPTSSTTSTISPAQK	VWVAPISGSVTTG	TKMWLTTKVGS	PATVTFQONKNFHQTATWVK 1980
Qy	1855	QGSNSGVVQVQKVLGII	IPSSGTSGOFTS	POPTATVTRPNTSGSGGTTNSOVIT	1914
Db	1981	QGSNSGVVQVQKVLGII	IPSSGTSGOFTS	POPTATVTRPNTSGSGGTTNSOVIT	2040
Qy	1915	GPOIRPGMTVIRTP	LQOQSTLGAII	IRTPVMP	QAPQVMTQIIRGQPVSTAVSAPNTVS 1974
Db	2041	GPOIRPGMTVIRTP	LQOQSTLGAII	IRTPVMP	QAPQVMTQIIRGQPVSTAVSAPNTVS 2100
Qy	1975	STPGQKSLTSATST	SNIOSSASQPPR	POQOVKLTMAOLT	TOLTOGHGNGOGLTVVLOGOG 2034
Db	2101	STPGQKSLTSATST	SNIOSSASQPPR	POQOVKLTMAOLT	TOLTOGHGNGOGLTVVLOGOG 2160
Qy	2035	QTTGQLQLIPQGVTVL	PGQOQLMQAAMP	NGTVQREFLTP	LATTAATATSTTTTSTTAA 2094
Db	2161	QTTGQLQLIPQGVTVL	PGQOQLMQAAMP	NGTVQREFLTP	LATTAATATSTTTTSTTAA 2220
Qy	2095	GTGEOROSKLSQPMQ	VHQDKTLP	PAOSSSVGPKAKA	QPTAQPSARQPTQOPQSPAQPEV 2154
Db	2221	GTGEOROSKLSQPMQ	VHQDKTLP	PAOSSSVGPKAKA	QPTAQPSARQPTQOPQSPAQPEV 2280
Qy	2155	QTOPEVQOTQTVSSHVP	SEAQPTHAOSSK	PQVAAQSQPQSNVQ	QSQSPVRVQSPQSTRIRP 2214
Db	2281	QTOPEVQOTQTVSSHVP	SEAQPTHAOSSK	PQVAAQSQPQSNVQ	QSQSPVRVQSPQSTRIRP 2340
Qy	2215	STPSQLSPGQOSQVQ	TTTSQPIPIQHTSLQ	IPSQGQPSQPOVQ	QSTQTLSSGOTLNQV 2274
Db	2341	STPSQLSPGQOSQVQ	TTTSQPIPIQHTSLQ	IPSQGQPSQPOVQ	QSTQTLSSGOTLNQV 2400
Qy	2275	SVSSPSRPOLOIQOP	QPOVIAVPOLOQOV	VLSQLSQSVVQ	IAQAOQSGVPOQIKLQLP 2334
Db	2401	SVSSPSRPOLOIQOP	QPOVIAVPOLOQOV	VLSQLSQSVVQ	IAQAOQSGVPOQIKLQLP 2460
Qy	2335	QIQOSSAVQTHQION	VYTVQAASVQEQ	LQVQLRDQOQK	KQOQIEIKREHTLOASNQS 2394
Db	2461	QIQOSSAVQTHQION	VYTVQAASVQEQ	LQVQLRDQOQK	KQOQIEIKREHTLOASNQS 2520
Qy	2395	EIIQOVVMKHNNAVIE	HLKOKKMTPAEREN	QRMVNCVMKYLDK	IDKEEKOAAKKR 2454
Db	2521	EIIQOVVMKHNNAVIE	HLKOKKMTPAEREN	QRMVNCVMKYLDK	IDKEEKOAAKKR 2580
Qy	2455	KREESVEQKSKONAT	KLISALLFKHKBOL	RAELKRALDKD	LQIEVQEBELKRLKIK 2514
Db	2581	KREESVEQKSKONAT	KLISALLFKHKBOL	RAELKRALDKD	LQIEVQEBELKRLKIK 2640
Qy	2515	EKDLMLQAQATAVA	APCPVTPVLP	PAPPPPPGVQHT	GLLSTPLPVASQKRKR 2574
Db	2641	EKDLMLQAQATAVA	APCPVTPVLP	PAPPPPPGVQHT	GLLSTPLPVASQKRKR 2700
Qy	2575	BEEDSSSKSKKKKMI	STTSKETKDKLYC	ICKTYPYDESKFY	IGCDRCQNMWHGRCVGI 2634
Db	2701	BEEDSSSKSKKKKMI	STTSKETKDKLYC	ICKTYPYDESKFY	IGCDRCQNMWHGRCVGI 2760

QY 2635 LOSEALIDEYVCPQCQSTEDAMVLTPLTEKDYGLKRVLSLQAHKMAWFLPEVDPN 2694
|||||
DB 2761 LOSEALIDEYVCPQCQSTEDAMVLTPLTEKDYGLKRVLSLQAHKMAWFLPEVDPN 2820
|||||
QY 2695 DAPDYGVVKEPMDLATMEERVORRYEKLTFEVDWTKIFDNCRYNPSDPFYOCAEV 2754
|||||
DB 2821 DAPDYGVVKEPMDLATMEERVORRYEKLTFEVDWTKIFDNCRYNPSDPFYOCAEV 2880
|||||
QY 2755 LESFFVQKLKGFKASRSHNNKLQSTAS 2781
|||||
DB 2881 LESFFVQKLKGFKASRSHNNKLQSTAS 2907
|||||
RESULT 3
ID AAO16418
XX AAO16418 standard; Protein; 2759 AA.
AC AAO16418;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein (NAAP) - SEQ ID No 15.
DE
XX
KW Human; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cancer; proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cell proliferative disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN W02003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US21179.
XX
PR 22-JUN-2001; 2001US-300518P.
PR 29-JUN-2001; 2001US-301787P.
PR 29-JUN-2001; 2001US-301792P.
PR 29-JUN-2001; 2001US-301892P.
PR 29-JUN-2001; 2001US-301893P.
PR 06-JUL-2001; 2001US-303405P.
PR 06-JUL-2001; 2001US-303442P.
PR 15-MAR-2002; 2002US-364438P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;
PI Borowsky ML, Yao MG, Wallia NK, Bandman O, Lal PG, Becha SD;
PI Lee SY, Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjadian Y;
PI Lu Y;
XX
DR WPI; 2003-201420/19.
DR N-PSDB; AAL51568.
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS)
XX
PS Claim 1; Page 234-241; 312pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia

CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC the invention are useful for creating transgenic animals to model human
CC disease. The present amino acid sequence represents a human nucleic acid-
CC associated protein of the invention.
XX
SQ Sequence 2759 AA:

Query Match 93.7%; Score 13428; DB 24; Length 2759;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 2625; Conservative 6; Mismatches 7; Indels 264; Gaps 3;

QY 1 MVSEEEEDGDAEETQDSEDEDEDEDEDDDDSDYPEEMEDDDDDASCTESSFRSHST 60
|||||
DB 1 MVSEEEEDGDAEETQDSEDEDEDEDEDDDDSDYPEEMEDDDDDASCTESSFRSHST 60
|||||
QY 61 YSSTPGRRKPRVHRPRSPILEEKDIPPLEFPKSSSEDLMPVNEHIMVIAIYVLRNFTV 120
|||||
DB 61 YSSTPGRRKPRVHRPRSPILEEKDIPPLEFPKSSSEDLMPVNEHIMVIAIYVLRNFTV 120
|||||
QY 121 LRLSPFRFDFCAALVSQEOCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
|||||
DB 121 LRLSPFRFDFCAALVSQEOCTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
|||||
QY 181 FIDGMTPEVLRVYCESDKKEYHHVLPYQAEADYPYGPVENKIKVLQFLVDQFLTTNIARE 240
|||||
DB 181 FIDGMTPEVLRVYCESDKKEYHHVLPYQAEADYPYGPVENKIKVLQFLVDQFLTTNIARE 240
|||||
QY 241 ELMSEGVIOYDHCRCVCHKLGLLCCETCSANYHLECVKPPLEVPEDQCEVCVAVKV 300
|||||
DB 241 ELMSEGVIOYDHCRCVCHKLGLLCCETCSANYHLECVKPPLEVPEDQCEVCVAVKV 300
|||||
QY 301 PGVTCVAEIQNKPYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENKKIWIYSTKVQ 360
|||||
DB 301 PGVTCVAEIQNKPYIRHEPIGYDRSRKRYWFLNRRLIIEEDTENENKKIWIYSTKVQ 360
|||||
QY 361 LAELIDCLDKYWEAELCKILEEMREEIHRHMDITEDLTNKARGSNKSFLLAANEELIES 420
|||||
DB 361 LAELIDCLDKYWEAELCKILEEMREEIHRHMDITEDLTNKARGSNKSFLLAANEELIES 420
|||||
QY 421 IRAKKGIDINVKSPETEEDKNETENDSKDAEKREFEFQDSLEKSDDKTTPDDPEQK 480
|||||
DB 421 IRAKKGIDINVKSPETEEDKNETENDSKDAEKREFEFQDSLEKSDDKTTPDDPEQK 480
|||||
QY 481 SE----- 482
||
DB 481 SEPTTEVGDGKNSVSNLGDNTTNATSETPSEGRSPVGCLETPDSSNMAEKKVASEL 540
|||||
QY 483 -----VGDFKSEKSNSESPGAGKSGSSTRIITRLRNPD SKLSQKSOQVAAAHEA 537
|||||
DB 541 PQDVPVGDGFKSEKSNSESPGAGKSGSSTRIITRLRNPD SKLSQKSOQVAAAHEA 600
|||||
QY 538 NKLFPKEGKEVLVYVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNQYSTNSPAL 597
|||||
DB 601 NKLFPKEGKEVLVYVNSQGEISRLSTKKEVIMKGNINNYFKLGQEGKYRVYHNQYSTNSPAL 660
|||||
QY 598 NKQHREDDHDKRRHLAHKFLTPAGFEKWNKSGVHSGSKVLTISTLRITITOLENNIPSSFL 657
|||||
DB 661 NKQHREDDHDKRRHLAHKFLTPAGFEKWNKSGVHSGSKVLTISTLRITITOLENNIPSSFL 720
|||||
QY 658 HPNWSHRANWIKAVQCMCKPREFALALAILCAVKPVVYMLPIWREFLGLTRLHRTSIE 717
|||||
DB 721 HPNWSHRANWIKAVQCMCKPREFALALAILCAVKPVVYMLPIWREFLGLTRLHRTSIE 780
|||||
QY 718 REEKEKVKKKKQBEETMQOATWVKYTFPVKHQVWKQKGEYRVGTGGGWSWISKTHV 777
|||||
DB 781 REEKEKVKKKKQBEETMQOATWVKYTFPVKHQVWKQKGEYRVGTGGGWSWISKTHV 840
|||||
QY 778 YRFVPLKPGTNNVYRKSLGKTKNNNDENDESCKRSRPPKIKIEPDSKDEKVGSD 837
|||||
DB 841 YRFVPLKPGTNNVYRKSLGKTKNNNDENDESCKRSRPPKIKIEPDSKDEKVGSD 900
|||||

Qy	838	AAKADQNMEDISKITEKDDQVKELDDSDSRCKEPEMEVDDDMKTESHVNCQESSQV	897
Db	901	AAKADQNMEDISKITEKDDQVKELDDSDSRCKEPEMEVDDDMKTESHVNCQESSQV	960
Qy	898	DVNVSGFHLRYSYKKTKSKLDGLLERIKQFTLEEKORLEBKILEGGIKIGIKTST	957
Db	961	DVNVSGFHLRYSYKKTKSKLDGLLERIKQFTLEEKORLEBKILEGGIKIGIKTST	1020
Qy	958	NSSKNLSESPVITKAKEGCOSDSMRQESPNANDOPEDLIQCSOSDSVLRMSDPSHT	1017
Db	1021	NSSKNLSESPVITKAKEGCOSDSMRQESPNANDOPEDLIQCSOSDSVLRMSDPSHT	1080
Qy	1018	TNKLYPRDRLDDVIRSPTKPKQNSIENDIEEKVSDLASRQEPKTSKTKGNDFFID	1077
Db	1081	TNKLYPRDRLDDVIRSPTKPKQNSIENDIEEKVSDLASRQEPKTSKTKGNDFFID	1140
Qy	1078	DSKIASADDTGLTICKNKKPLIOESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMDFE	1137
Db	1141	DSKIASADDTGLTICKNKKPLIOESDTIVSSSKSALHSSVPKSTNDRDATPLSRAMDFE	1200
Qy	1138	GKLGCDSESNSTLNSDTSVIOSSSEEDMIVQNSNESISEQFTRRDQVEVLEPLKCEL	1197
Db	1201	GKLGCDSESNSTLNSDTSVIOSSSEEDMIVQNSNESISEQFTRRDQVEVLEPLKCEL	1260
Qy	1198	VSGESTGNCEDRLPVKGTANGKPKSOQKLEERPYNKCSQIKLKNTTDKKNNENRESE	1257
Db	1261	VSGESTGNCEDRLPVKGTANGKPKSOQKLEERPYNKCSQIKLKNTTDKKNNENRESE	1320
Qy	1258	KKGQRTSTFOINGDKNPKIYLKGECLKEISESRVSGNVEPKVNNINKLIIPENDIKSLT	1317
Db	1321	KKGQRTSTFOINGDKNPKIYLKGECLKEISESRVSGNVEPKVNNINKLIIPENDIKSLT	1380
Qy	1318	VKESAIRPFINGDVIEMDFERNESSETKSHLLSSSDAEGNRYDSLETLPSTKESDSTQTT	1377
Db	1381	VKESAIRPFINGDVIEMDFERNESSETKSHLLSSSDAEGNRYDSLETLPSTKESDSTQTT	1440
Qy	1378	TPSASCPESNSVQVEDMEITSEVKVTSPTSEESNLNSDFIDENGLPIPNKNENVN	1437
Db	1441	TPSASCPESNSVQVEDMEITSEVKVTSPTSEESNLNSDFIDENGLPIPNKNENVN	1500
Qy	1438	GESRKKVITETVTMTSTVATESKTVIKVEGDKQTVVSSSTENCAKSTVTTTTVTVKLS	1497
Db	1501	GESRKKVITETVTMTSTVATESKTVIKVEGDKQTVVSSSTENCAKSTVTTTTVTVKLS	1560
Qy	1498	TPSTGGSDIISVKEQSKTVVTTVTDLSLTGTGLTVTSMTVSKEYSTRDKVKLMPFSRP	1557
Db	1561	TPSTGGSDIISVKEQSKTVVTTVTDLSLTGTGLTVTSMTVSKEYSTRDKVKLMPFSRP	1620
Qy	1558	KKTRSGTALPSYRKFTVKTSKSIIFVLPNDLKLARKGGIREVPYFNYNAPALDIWPY	1617
Db	1621	KKTRSGTALPSYRKFTVKTSKSIIFVLPNDLKLARKGGIREVPYFNYNAPALDIWPY	1680
Qy	1618	PSPRPTFGITWRYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGGSGTTRTSETEI	1677
Db	1681	PSPRPTFGITWRYRLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGGSGTTRTSETEI	1740
Qy	1678	TTTEIIKRRDVGPIREYECIRKIICIGVPETPKETPPQKGLRSSALRPRKRPETPK	1737
Db	1741	TTTEIIKRRDVGPIREYECIRKIICIGVPETPKETPPQKGLRSSALRPRKRPETPK	1800
Qy	1738	QTGPVITETWAAEELELEWIRAFARVEKEKAQAVEQAKRLEQOQKPTVIATSTTSPT	1797
Db	1801	QTGPVITETWAAEELELEWIRAFARVEKEKAQAVEQAKRLEQOQKPTVIATSTTSPT	1860
Qy	1798	SSTTSTISPAQKVMVAPISGVSSTGKMLVLTTKVGSPTATVFOQKNKPHOTFATVWKQOQ	1857
Db	1861	SSTTSTISPAQKVMVAPISGVSSTGKMLVLTTKVGSPTATVFOQKNKPHOTFATVWKQOQ	1920
Qy	1858	SNSGVVQVQKVLGIIPSTGTSQOFTTSFQPRATVTIIRNPTSGSGGTTNSQVITGPQ	1917
Db	1921	SNSGVVQVQKVLGIIPSTGTSQOFTTSFQPRATVTIIRNPTSGSGGTTNSQVITGPQ	1980
Qy	1918	IRPGMTVIRTPLQQSTGLKAIIRTPVMVPGAPQVMTQIIRGQPVSTAVSAPNTVSSSTP	1977

Db	1981	IRPGMTVIRTPLQQSTGLKAIIRTPVMVPGAPQVMTQIIRGQPVSTAVSAPNTVSSSTP	2040
Qy	1978	GOKSLTSATSTSNIQSSASOPPRPOQOVKLTMAQLTQLTQHGNGOGLTVVITQOGQTT	2037
Db	2041	GOKSLTSATSTSNIQSSASOPPRPOQOVKLTMAQLTQLTQHGNGOGLTVVITQOGQTT	2100
Qy	2038	GQLQLIPQGVTVLPFGQOQLMAAMPNGTVQRFLFTPLATTATTASTTTTSTVSTAAGTG	2097
Db	2101	GQLQLIPQGVTVLPFGQOQLMAAMPNGTVQRFLFTPLATTATTASTTTTSTVSTAAGTG	2160
Qy	2098	EQRSKLSFOMQVHQDKTLPPAQSSVSGAKAQPOTAQPSARPOQOTOPQSPQAQPEVQIQ	2157
Db	2161	EQRSKLSFOMQVHQDKTLPPAQSSVSGAKAQPOTAQPSARPOQOTOPQSPQAQPEVQIQ	2220
Qy	2158	PEVQOTVTVSSHVSPSEAQPTHQAQSSKQVAAQSQPSQVQSGSPVVRVQSPSOTRIRPSTP	2217
Db	2221	PEVQOTVTVSSHVSPSEAQPTHQAQSSKQVAAQSQPSQVQSGSPVVRVQSPSOTRIRPSTP	2280
Qy	2218	SQSPGQOQSOVQTTTSTQPIQPIHTSLQIPSGQOQSOQOVQVQSTQTLSSGGQTLNQVSVS	2277
Db	2281	SQSPGQOQSOVQTTTSTQPIQPIHTSLQIPSGQOQSOQOVQVQSTQTLSSGGQTLNQVSVS	2319
Qy	2278	SPSRPQLOIQOPQOVYLAQVQLQOQVQLSQIQSVVAQIQAAQSGVPPQIKLQLPIQIQ	2337
Db	2320	-----	2319
Qy	2338	QSSAVQTHQTVVTVQAASVQEQLRVQOQRDQOQKKQOQIEIKREHTLQASNOSEII	2397
Db	2320	-----	2319
Qy	2398	QKQVVMKHNNAVIEHLKQKSKMTFAERENORMIVQVQVMIYILDKIDKEKQAAKKRKE	2457
Db	2320	--QVVMKHNNAVIEHLKQKSKMTFAERENORMIVQVQVMIYILDKIDKEKQAAKKRKE	2377
Qy	2458	ESVEQKRSKONATKLSALLFKHKEQLRAEILKKRALLDKDLQIEVQBELKRDLIKKEKD	2517
Db	2378	ESVEQKRSKONATKLSALLFKHKEQLRAEILKKRALLDKDLQIEVQBELKRDLIKKEKD	2437
Qy	2518	LMQLAQATAVAAPCPVPTVLPAPPAPPPPPPPGVQHTGGLLSTPTPLPVAQSKRKEEE	2577
Db	2438	LMQLAQATAVAAPCPVPTVLPAPPAPPPPPPPGVQHTGGLLSTPTPLPVAQSKRKEEE	2497
Qy	2578	KDSSSKSKKKMISTTTSKETKDTKLYCICKTPYDESK-----FYIGCDRCQNNYHGRGVGILQSEA	2615
Db	2498	KDSSSKSKKKMISTTTSKETKDTKLYCICKTPYDESKFYIGCDRCQNNYHGRGVGILQSEA	2557
Qy	2616	-----FYIGCDRCQNNYHGRGVGILQSEA	2639
Db	2558	EAKKMDVYICNDCKRAQEGSSEELYCICRTPTDESQFYIGCDRCQNNYHGRGVGILQSEA	2617
Qy	2640	ELIDEEYVCPQOQSTEDAMTVLTPLEKDYEGELKRVLSLQAHKMAWPFLEVPDNDAPDY	2699
Db	2618	ELIDEEYVCPQOQSTEDAMTVLTPLEKDYEGELKRVLSLQAHKMAWPFLEVPDNDAPDY	2677
Qy	2700	YGVITKEPMDLATMEERVORRYEKLTEFVADMTKIFONCRYPNPSDPFFYQCAEVLSEFF	2759
Db	2678	YGVITKEPMDLATMEERVORRYEKLTEFVADMTKIFONCRYPNPSDPFFYQCAEVLSEFF	2737
Qy	2760	VQKLGKFKASRSHNNKLQSTAS	2781
Db	2738	VQKLGKFKASRSHNNKLQSTAS	2759

RESULT 4
AAB94078
ID AAB94078 standard; Protein; 557 AA.
XX
AC AAB94078;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14273.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 XX PR 29-JUL-1999; 99JP-0248036.
 XX PR 27-AUG-1999; 99JP-0300253.
 XX PR 11-JAN-2000; 2000JP-0118776.
 XX PR 02-MAY-2000; 2000JP-0183767.
 XX PR 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX DR WPI: 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX PS Claim 8; SEQ ID 14273; 2537pp + CD ROM; English.
 XX CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC polynucleotide and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX SQ Sequence 557 AA;
 Query Match 18.8%; Score 2695.5; DB 22; Length 557;
 Best Local Similarity 79.0%; Pred. No. 1.7e-130;
 Matches 553; Conservative 3; Mismatches 1; Indels 143; Gaps 1;
 QY 1811 WYAPISGSVTTGKMWLTKVGSPTATVTFTQKNKHFQFATVWKQGQNSGVVQVQKVL 1870
 Db 1 MVAPISGSVTTGKMWLTKVGSPTATVTFTQKNKSFQFATVWKQGQNSGVVQVQKVL 60
 QY 1871 GIIPSTGTSQQTFTSFQRTATVTIRPNTSGSGTTSNSQVITGPQIRPGMTVIRTPQ 1930
 Db 61 GIIPSTGTSQQTFTSFQRTATVTIRPNTSGSGTTSNSQVITGPQIRPGMTVIRTPQ 120
 QY 1931 QSTLGAIIIRTPVWQAPQVMTQIIRGQSPVSTAVSAPNTVSTPGQKSLTSATSTSN 1990
 Db 121 QSTLGAIIIRTPVWQAPQVMTQIIRGQSPVSTAVSAPNTVSTPGQKSLTSATSN 180
 QY 1991 IOSSASQPPRPOGGQVKLTMAQLTQLTQGHGNGQLTVVIOGGQTTGQLQLIPQGVTVL 2050

Db 181 IOSSASQPPRPOGGQVKLTMAQLTQLTQGHGNGQLTVVIOGGQTTGQLQLIPQGVTVL 240
 QY 2051 PPGGQOLMQAAMPNGTVQRFLEFTPLATTATTATTTTSTTAAGTGEORQSKLSQMOV 2110
 Db 241 PPGGQOLMQAAMPNGTVQRFLEFTPLATTATTATTTTSTTAAGTGEORQSKLSQMOV 300
 QY 2111 HQDKTLPPAQSSTSSVGPAAQPAQPAQPAQPAQPAQPAQPAQPAQPAQPAQPAQPAQ 2170
 Db 301 HQDKTLPPAQSSTSSVGPAAQPAQPAQPAQPAQPAQPAQPAQPAQPAQPAQPAQPAQ 360
 QY 2171 PSEAQPTHQAQSSKPPQVAAQSQPSNVQSGSPVRVQSPQSPQSPQSPQSPQSPQSPQ 2230
 Db 361 PSEAQPTHQAQSSKPPQVAAQSQPSNVQSGSPVRVQSPQSPQSPQSPQSPQSPQSPQ 420
 QY 2231 TTSQPIPIQPIHTSLQIPSGQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQ 2290
 Db 421 TTSQPIPIQPIHTSLQIPSGQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQ 446
 QY 2291 PQVIAVPQLQQVQVLSQISQSVVQIAQISQSVVQIAQISQSVVQIAQISQSVVQIAQIS 2350
 Db 447 ----- 446
 QY 2351 VTVQASVQEQQLQVQQLRDQDQKKQKKQKKQKKQKKQKKQKKQKKQKKQKKQKKQKK 2410
 Db 447 -----QVVMKHNNAVIE 457
 QY 2411 HLKQKSMTPAEREENORMIVCNQVMKYITLTKIDKEKQAAKRRKREESVEQKRKONAT 2470
 Db 458 HLKQKSMTPAEREENORMIVCNQVMKYITLTKIDKEKQAAKRRKREESVEQKRKONAT 517
 QY 2471 KLSALLFKHKEQLRAEILKRALDKDLQIEVQEEELKRD 2510
 Db 518 KLSALLFKHKEQLRAEILKRALDKDLQIEVQEEELKRD 557
 RESULT 5
 ID ABB68735
 ID ABB68735 standard; Protein; 976 AA.
 AC ABB68735;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 32997.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL12838.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 32997; 21pp + Sequence Listing; English.
 XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB157737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 976 AA;

Query Match 10.9%; Score 1557.5; DB 22; Length 976;
Best Local Similarity 36.48; Pred. No. 1.4e-71;
Matches 357; Conservative 138; Mismatches 222; Indels 193; Gaps 22;
QY 4 EEEEEEDGAEETQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESFRSHSTYSS 63
DB 99 ESEHYGSDGDFGSEEDKSDNEDMLTPSDDESLEAVANESSEFSVC---SFGNGV--- 152
QY 64 TPGRRKPRVHRPRSPILAE-KDIPPLEFPKSSDLMVPNEHIMNVIAYIEVLRNFGTVLR 122
DB 153 ---GRPRPPSPPEVWLOEGRQYAAALDLPDSSDLFIANTHVLRALSIYEVLRFRHMVR 209
QY 123 LSPFRFDFCAALYSQEOCTLMAEMHVLLKAVLREEDTNTVFGPADLKDSVNSTLYFI 182
DB 210 LSPFRFDFCAALACEQSALLTEVHIMLLKAILREEDAQGTGPDLDQDVTNLSLYLI 269
QY 183 DGMTWPEVLRVYCSDEKEY-----HHVLPYQAEADYPYGPVENKTKVQLFLVDQLTTNIA 238
DB 270 DSITWPEVLRYSVESDKTFDRNVRPHILSHTF---YPYTGDINDRLEVQLFSDQPLTNSI 326
QY 239 REELMSEGVIOYDCHRVCHKGLDLCCTCSAYVHLECVKPLPEEVEDWQCEVCVAH 298
DB 327 RDVLMQEGPIHIDCHRVCHRLGLDLCCTCPAVYHLECVDPDMDVPTDQWGLGRSH 386
QY 299 KVPGVTCVARTQKNPKYRHERPGYCDYRKRKYFLNRLTIETDENENEKKIWIYSTK 358
DB 367 KVSQVDCVLPQEGVLIHSDSLGVDHGRKYFIARRIFI-EDQENFT---CWYISTT 442
QY 359 VOLAELIDCLDKYWEALCKILEEMRETHRHMDITDLTNKARGNSKSFLLAANEIL 418
DB 443 SKLLLSRLDAELETFLHSQITERRDEIERQMKLTETLTNEHKHTKRSVI----- 494
QY 419 ESIRAKKGIDNVKSPRETKDNKTENDSKDAEKNREEDQSLKSDDKTPDDDPQ 478
DB 495 -----EIQEAKN-----ELLEKEVLDEDEKD----- 516
QY 479 GKSEVGFDPKSKNGELSESPGAGGAGSTRIITRLRNPDSKLSQKQVAAAHEAN 538
DB 517 -----GDAKSESQSTF-----GTKKECKMVT-----QKSNLT----- 547
QY 539 KLFKEGKVLVNVSGEISRLSTKKEVIMKNINNYFKLGOEGKYRVYHNOYSTNSPALN 598
DB 548 -----NGTLHFKLGMEOGFKNVYNOYSTNPALN 576
QY 599 KQHREDDKRRHLAKFCLTPAGEFKWNGSVHSGKVLTIISTRLTITQLENNIPSSPLH 658
DB 577 KQORNEEDKRRHLHSHKSLTASDFKIGITGMTDNTMTLTLQTLINFESNIAAFLN 636
QY 659 PNWASHRANWIKAVOMSKSPREFALALALECAVKPVVMLPIWREFLGHTRHRTWSTIER 718
DB 637 INWVVKKIWNAAVNARNRPSSEFAVLLLFQASLKSVMFANVWHEQLGHTTLQRTISAER 696
QY 719 BEKEVKKKKKQ-----EEETMQQATWVKYFPVKHVKOKGGEYRVTVGGYGSWISKT 775
DB 697 EERKLEKREKREDEERLRAFNYIKYTLGLKHQVWKOKGEYRVHVGQWGLWLSSS 756
QY 776 -----HVYRFPKLPGNVNVYRKSLEGTKNNDMDSDSKRCSRSPKKIKIPDSEK 830
DB 757 RCGVRRARQPLTHNVVYVHTM---GEENDVNEII-----LVDPRQRFMQCES 805

QY 831 DEYKGSAAKAGDQ---NEMDISKITEKKDQDVKELLDSDSKPCKEPMEDDMKTESH 888
DB 806 SNVDGQGVCHLPQYKNVKVIEDVTE-----KIKGH 836
QY 889 VNCQESSQDVVNVSEGFHL--RTSYKKKTKSSKLDGGLLERRIKQFTLEEKQRLKIKLE 946
DB 837 -----IDVSKALNAPGRTYYSKVARKSLDLDLDRRLKLAEEVQM---ASKIP 882
QY 947 GGIKIGIKTSTNSKNLSES 966
DB 883 SDMKPLLVSSQNNTANSKQT 902

RESULT 6
AAB54364
ID AAB54364 standard; Protein; 328 AA.
XX
AC AAB54364;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:816.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX
OS Homo sapiens.
XX
PN WO2000055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-579444/54.
DR N-PSDB; AAC99129.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 11; Page 1272-1273; 1379pp; English.

XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The

CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 328 AA;

Query Match 10.3%; Score 1474; DB 21; Length 328;
Best Local Similarity 94.8%; Pred. No. 6.1e-68;
Matches 289; Conservative 0; Mismatches 14; Indels 2; Gaps 1;

QY 406 NKFLAAANEILIESIRAKGDDIDNVKSP--EETEDKNETENDSKDAENREFEQSL 463

Db 12 NVSLSGGTNEILIESIRAKGDDIDNVKSPGTEETEDKNETENDSKDAENREFEQSL 71

QY 464 EKSDDKTPDDPEQKSEVGFDFKSEKNGELSESPGAGSGSTRITRLRNPDSKLS 523

Db 72 EKSDDKTPDDPEQKSEVGFDFKSEKNGELSESPGAGSGSTRITRLRNPDSKLS 131

QY 524 QLKSOQVAAAHAANKLFEGKEVLVVVNSQGEISRLSTKKEVTMKGNINNYFKLGQEGY 593

Db 132 QLKSOQVAAAHAANKLFEGKEVLVVVNSQGEISRLSTKXXVXMKGXINNYXKLQEGY 191

QY 584 RYVHNOYSTNSPALNKHQREDHDKRHLAKFCLTPAGEFKWNGSVHGSKVLTITLRL 643

Db 192 RYVHNOYSTNSPALNKHQREDHDKRHLAKFCLTPAGEFKWNGSVHGSKVLTITLRL 251

QY 644 TITOLENNIPSSPLHPNASHRANWIKAVQMCSPREFALALAILCAVKKPVVYMLPIWRE 703

Db 252 TITOLENNIPSSPLHPNASHRANWIKAVQMCSPREFALALAILCAVKKPVVYMLPIWRE 311

QY 704 FLIGHT 708

Db 312 SLIGHT 316

RESULT 7

ABU70846

ID ABU70846 standard; Protein; 191 AA.

XX AC ABU70846;

DT 10-JUN-2003 (first entry)

XX Human adipocyte Selected Interacting domain, SID, #477.

DE Human; prev; adipocyte; SID; selected interacting domain;

KW anorectic; antidiabetic; protein-protein interaction; diabetes;

KW yeast 2-hybrid assay; metabolic disorder; obesity.

XX OS Homo sapiens.

XX WO200286122-A2.

XX 31-OCT-2002.

XX 14-MAR-2002; 2002WO-EP03768.

XX 14-MAR-2001; 2001US-275734P.

XX (HYBR-) HYBRIGENICS.

XX Legrain P, Daviet L;

XX WPI; 2003-103412/09.

XX N-PSDB; ACA57390.

XX New complex between two interacting proteins in adipocyte cells, useful

PT for identifying selected interacting domains that modulate protein

PT interactions, or for preventing or treating metabolic disorders such as

PT obesity or diabetes

XX Claim 6; Page 267-268; 382pp; English.

XX The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC polynucleotide encoding a polypeptide in the adipocyte cells, a
CC recombinant host cell expressing at least one of the interacting
CC polypeptides of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
CC sequences given in the specification (including its fragment or variant),
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC given in the specification (including its fragment or variant), a vector
CC comprising the SID (RTM) polynucleotide, a recombinant host cell
CC comprising the vector, a protein chip comprising the polypeptides and
CC a record comprising all or part of the data, listed in the specification.
CC The complex, polypeptides, polynucleotides and compounds are
CC useful for preventing or treating metabolic disorders such as obesity
CC or diabetes. The polynucleotides are useful as probes or primers. The
CC complex is particularly useful for identifying selected interacting
CC domains (SID (RTM)) for screening drugs that modulate the protein
CC interaction, thus exhibiting the therapeutic effect. The present
CC sequence represents a SID (prey) protein of the invention.

XX Sequence 191 AA;

Query Match 7.2%; Score 1030; DB 24; Length 191;

Best Local Similarity 100.0%; Pred. No. 2e-45; 0; Gaps 0;

Matches 191; Conservative 0; Mismatches 0; Indels 0;

QY 2556 HTGLLSTPTLPVASKRKRREKSSSKKKKKMISTTSKTKKDKLYCICKTPYDESK 2615

Db 1 HTGLLSTPTLPVASKRKRREKSSSKKKKKMISTTSKTKKDKLYCICKTPYDESK 60

QY 2616 FYIGCDRCQNWYHRCVGLQSEAEILDEYVCPQCSTEDAMTVLPLEKDEGLKRVL 2675

Db 61 FYIGCDRCQNWYHRCVGLQSEAEILDEYVCPQCSTEDAMTVLPLEKDEGLKRVL 120

QY 2676 RSLQAHKMAWPFLEPVDNDADPDYGVYKPEMDLATMEERVQRRYKLETFEVDMTKIF 2735

Db 121 RSLQAHKMAWPFLEPVDNDADPDYGVYKPEMDLATMEERVQRRYKLETFEVDMTKIF 180

QY 2736 DNCRYNPSDS 2746

Db 181 DNCRYNPSDS 191

RESULT 8

AAU04323

ID AAU04323 standard; Protein; 238 AA.

XX AC AAU04323;

XX 18-JUN-1999 (first entry)

DT Human secreted protein SEQ ID NO:69.

DE Human; secreted protein; cancer; tumour; developmental abnormality;

KW foetal deficiency; blood disorder; immune system disorder; inflammation;

KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;

KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;

KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;

KW digestive disorder; endocrine disorder; infection; AIDS.

XX OS Homo sapiens.

XX WO9910363-A1.

XX 04-MAR-1999.

XX 27-AUG-1998; 98WO-US17709.

XX

```
PR 29-AUG-1997; 97US-0056271.
PR 29-AUG-1997; 97US-0056073.
PR 29-AUG-1997; 97US-0056247.
PR 29-AUG-1997; 97US-0056270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
XX
XX WPI; 1999-190585/16.
XX
XX New isolated human genes and the secreted polypeptides they encode
XX - useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Disclosure; Page 18; 170pp; English.
XX
XX AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.
XX AAY04293 to AAY04321 represent the secreted proteins encoded by the 29
XX human genes. The genes and their corresponding secreted polypeptides are
XX useful for preventing, treating or ameliorating medical conditions,
XX e.g. by protein or gene therapy. Also pathological conditions can be
XX diagnosed by determining the amount of the new polypeptides in a sample
XX or by determining the presence of mutations in the new genes. Specific
XX uses are described for each of the 29 genes, based on which tissues they
XX are most highly expressed in, and include developing products for the
XX diagnosis or treatment of cancer, tumours, developmental abnormalities
XX and foetal deficiencies, blood disorders, diseases of the immune system,
XX autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
XX disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
XX disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
XX disorders, digestive/endocrine disorders, infections and AIDS. The
XX polypeptides are also useful for identifying their binding partners.
XX The sequences given in AAX30174 to AAX30182 and AAY04322 to AAY04334 are
XX used in the exemplification of the present invention.
XX
XX Sequence 238 AA;
XX
XX Query Match 6.9%; Score 982.5; DB 20; Length 238;
XX Best Local Similarity 78.8%; Pred. No. 7.8e-43;
XX Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;
XX
QY 2563 PTLVASOKRKEEK-----DSSSKSKKKMISTTSKYNK-----DT 2601
DQ 3 PTPR----KTPYDESKFYIGCDLCTNWHGCVGITEKAKKMDVYICNCKKRAQGSSE 58
QY 2602 KLYCICKTPYDESKFYIGCDRCQNWYHGRVCVGIQSEAELIDVYVQCQSTEDAMTVLT 2661
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 59 ELYCICRTPYDESQFYIGCDRCQNWYHGRVCVGIQSEAELIDVYVQCQSTEDAMTVLT 118
QY 2662 PLTEKDYEGLRVLRSLQAKHMAWPFLEPVPDNDAPDYGYVGIKEPMDLATMEERVQRRY 2721
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 119 PLTEKDYEGLRVLRSLQAKHMAWPFLEPVPDNDAPDYGYVGIKEPMDLATMEERVQRRY 178
QY 2722 EKLTFEVDAMTKIFDNCRYNPSDPFYQCAEVLESFPVQKLGFKASRSNNKLOSTAS 2781
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 179 EKLTFEVDAMTKIFDNCRYNPSDPFYQCAEVLESFPVQKLGFKASRSNNKLOSTAS 238
XX
RESULT 9
AAU19446
ID AAU19446 standard; Protein; 237 AA.
XX
XX AAU19446;
XX
XX 04-DEC-2001 (first entry)
XX
XX Human diagnostic and therapeutic polypeptide (DITHP) #32.
XX
XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
XX cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
XX acquired immune deficiency syndrome; AIDS; autoimmune disorder;
XX respiratory disorder.
```

```
XX OS Homo sapiens.
XX PN WO200162927-A2.
XX PD
XX PD 30-AUG-2001.
XX
XX 21-FEB-2001; 2001WO-US06059.
XX
XX 24-FEB-2000; 2000US-0184693.
XX 24-FEB-2000; 2000US-0184697.
XX 24-FEB-2000; 2000US-0184698.
XX 24-FEB-2000; 2000US-0184768.
XX 24-FEB-2000; 2000US-0184769.
XX 24-FEB-2000; 2000US-0184770.
XX 24-FEB-2000; 2000US-0184771.
XX 24-FEB-2000; 2000US-0184772.
XX 24-FEB-2000; 2000US-0184773.
XX 24-FEB-2000; 2000US-0184774.
XX 24-FEB-2000; 2000US-0184776.
XX 24-FEB-2000; 2000US-0184777.
XX 24-FEB-2000; 2000US-0184797.
XX 24-FEB-2000; 2000US-0184813.
XX 24-FEB-2000; 2000US-0184837.
XX 24-FEB-2000; 2000US-0184841.
XX 24-FEB-2000; 2000US-0185213.
XX 24-FEB-2000; 2000US-0185216.
XX 12-MAY-2000; 2000US-0203785.
XX 15-MAY-2000; 2000US-0204226.
XX 16-MAY-2000; 2000US-0204525.
XX 16-MAY-2000; 2000US-0204821.
XX 16-MAY-2000; 2000US-0204908.
XX 16-MAY-2000; 2000US-0205232.
XX 17-MAY-2000; 2000US-0204815.
XX 17-MAY-2000; 2000US-0204863.
XX 17-MAY-2000; 2000US-0205221.
XX 17-MAY-2000; 2000US-0205285.
XX 17-MAY-2000; 2000US-0205286.
XX 17-MAY-2000; 2000US-0205287.
XX 17-MAY-2000; 2000US-0205323.
XX 17-MAY-2000; 2000US-0205324.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
XX Chen A, D'sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE;
XX DuFour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
XX Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
XX Wright RJ, Yap PE, Yu JW, Bradley DL, Bratcher SR, Chen W;
XX Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
XX WPI; 2001-502867/55.
XX N-PSDB; AAS31017.
XX
XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
XX enzymes, hormones and receptors, useful in diagnostics and therapeutics
XX
XX Claim 27; Page 418-419; 522pp; English.
XX
XX The invention relates to polynucleotides (I) encoding diagnostic and
XX therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
XX and proteins involved in growth and development and receptors. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate DITHP expression. For example, (I) and
XX (II) may be used to treat disorders associated with decreased polypeptide
XX expression by rectifying mutations or deletions in a patient's genome,
XX that affect the activity of the DITHPs, by expressing inactive proteins
XX or supplementing the patient's own production of them. (I) and (II)
XX may be used to treat diseases, for example, cell proliferative disorder,
XX Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
XX leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
XX (I) may be used to produce the DITHPs, by inserting the nucleic acids
```

CC into a host cell and culturing the cell to express the protein. (I) and
 CC its complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. (II) may also be used as antigens in the production of
 CC antibodies against DTHPs and in assays to identify modulators of DTHP
 CC expression and activity. The anti-DTHP antibodies and antagonists may
 CC also be used to down regulate expression and activity. The anti-DTHP
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of DTHPs in samples (e.g. by enzyme linked immunosorbant
 CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and
 CC therapeutic (DTHP) polypeptides of the invention.

XX Sequence 237 AA;
 SQ
 Query Match 6.8%; Score 981; DB 22; Length 237;
 Best Local Similarity 87.5%; Pred. No. 9.2e-43;
 Matches 182; Conservative 4; Mismatches 6; Indels 16; Gaps 1;

QY 2590 ISTSTETKK-----DTKLYICKTPYDESKFYIGCDRCQNNYHGRGVG 2633
 DB 30 VGITEKAKKMDYIICNDCKRAQGSSEELYICITPYDESQFYIGCDRCQNNYHGRGVG 89

QY 2634 ILQSEALIDEYVCPQOSTEDAMTVLPTLTKDYEGKLRVLSLQAHKMAWPFLEPVD 2693
 DB 90 ILQSEALIDEYVCPQOSTEDAMTVLPTLTKDYEGKLRVLSLQAHKMAWPFLEPVD 149

QY 2694 NDAPDYGVYKPEMDLATMEERVQRRYVEKLTFFVADMTKIFDNCRYNPSDSPFFYQCAE 2753
 DB 150 NDAPDYGVYKPEMDLATMEERVQRRYVEKLTFFVADMTKIFDNCRYNPSDSPFFYQCAE 209

QY 2754 VLESFFVQKLGFKASRSHNNKLOSTAS 2781
 DB 210 VLESFFVQKLGFKASRSHNNKLOSTAS 237

RESULT 10
 AAU78845
 ID AAU78845 standard; Protein; 149 AA.
 XX AC AAU78845;
 DT 18-JUN-2002 (first entry)
 XX DE SCAN/KRAB protein SKAT-2 binding protein, 7A1.
 XX
 KW SCAN/KRAB protein associated with a Th2 phenotype; SKAT-2; asthma;
 KW allergy; hay fever; atopic dermatitis; allergic rhinitis;
 KW Alzheimer's disease; neurological disorder; apoptosis;
 KW T-helper response; vaccination; immune response; autoimmune disease;
 KW infection; cancer; kruppel-like zinc finger protein; 7A1; human.
 XX OS Homo sapiens.
 XX GB2364051-A.
 XX 16-JAN-2002.
 XX 06-APR-2000; 2000GB-0008549.
 XX 06-APR-2000; 2000GB-0008549.
 XX (GLAXO) GLAXO GROUP LTD.
 XX Blanchard AD, Champion B, Page KR;
 XX WPI; 2002-306794/35.
 DR N-PSDB; ABK47549.
 XX
 XX New SKAT-2 polypeptide transcription factor, useful for treating or
 PT preventing e.g. allergy and Alzheimer's disease, also related nucleic
 PT acid, antibodies and modulators -
 XX

PS Disclosure; Page 37-38; 47pp; English.
 XX
 CC The invention describes an isolated SCAN/KRAB protein associated with a
 CC Th2 phenotype (SKAT-2) polypeptide (I). (I), especially where expressed
 CC in recombinant cells, is used to identify agents that modulate SKAT-2
 CC activity, and may be useful therapeutically. The agents, also (I) and
 CC the polynucleotides (II) that encode it, are useful for treatment and
 CC prevention of diseases responsive to SKAT-2 modulation, specifically
 CC asthma; allergies (hay fever, atopic dermatitis; allergic rhinitis etc.);
 CC Alzheimer's disease (and other neurological disorders, particularly where
 CC related to apoptosis); conditions involving T-helper responses, and
 CC apoptosis; also in vaccination against an antigen to enhance/repress Th2
 CC or humoral responses, to reduce development of a Th1 phenotype, and to
 CC manipulate the immune response in autoimmune diseases, infections and
 CC cancer. (II) is also used for recombinant production of (I); as source
 CC of antisense therapeutics and as primers and probes, e.g. for diagnostic
 CC detection of mutations and for monitoring SKAT-2 expression in
 CC association with disease. This is the amino acid sequence of the peptide
 CC 7A1, that binds specifically to the kruppel-like zinc finger protein,
 CC SKAT-2.
 XX
 SQ Sequence 149 AA;
 Query Match 5.5%; Score 789; DB 23; Length 149;
 Best Local Similarity 100.0%; Pred. No. 3.7e-33;
 Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2633 GILQSEALIDEYVCPQOSTEDAMTVLPTLTKDYEGKLRVLSLQAHKMAWPFLEPVD 2692
 DB 1 GILQSEALIDEYVCPQOSTEDAMTVLPTLTKDYEGKLRVLSLQAHKMAWPFLEPVD 60

QY 2693 PDAPDYGVYKPEMDLATMEERVQRRYVEKLTFFVADMTKIFDNCRYNPSDSPFFYQCA 2752
 DB 61 PDAPDYGVYKPEMDLATMEERVQRRYVEKLTFFVADMTKIFDNCRYNPSDSPFFYQCA 120

QY 2753 EVLESFFVQKLGFKASRSHNNKLOSTAS 2781
 DB 121 EVLESFFVQKLGFKASRSHNNKLOSTAS 149

RESULT 11
 AAU31944
 ID AAU31944 standard; Protein; 246 AA.
 XX AC AAU31944;
 XX 18-DEC-2001 (first entry)
 XX DE Novel human secreted protein #2435.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX OS Homo sapiens.
 XX WO200179449-A2.
 XX 25-OCT-2001.
 XX 16-APR-2001; 2001WO-US08656.
 XX 18-APR-2000; 2000US-0552929.
 XX 26-JAN-2001; 2001US-0770160.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT;
 XX WPI; 2001-611725/70.
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 PT

XX PS Claim 20; Page 529-530; 765pp; English.

XX CC The invention relates to novel human secreted polypeptides. The

CC polypeptides and antibodies to the polypeptides are useful for

CC determining the presence of or predisposition to a disease associated

CC with altered levels of polypeptide. The polypeptides are also useful for

CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent

CC for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising

CC the nucleic acids encoding the polypeptides and cells genetically

CC engineered to express them are also useful for producing the proteins.

CC The proteins are useful in genetic vaccination, testing and

CC therapy, and can be used as nutritional supplements. They may be used to

CC increase stem cell proliferation; to regulate haematopoiesis; and in

CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;

CC immune suppression and/or stimulation; as anti-inflammatory agents; and

CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid

CC sequences of novel human secreted proteins of the invention.

XX SQ Sequence 246 AA;

Query Match 5.4%; Score 781; DB 22; Length 246;

Best Local Similarity 67.8%; Pred. No. 1.9e-32;

Matches 162; Conservative 10; Mismatches 63; Indels 4; Gaps 2;

QY 751 HOVWKQGEYRVNIGYGSWISKTHYRVFKPLPGNTNINRKSLGRTKNNMDENDES 810

DB 7 HOVWKQGEYRVNIGYGSWISKTHYRVFKPLPGNTNINRKSLGRTKNNMDENDES 66

QY 811 DKRCRSRPPKIKIEPDSEKDEKVGSDAAGADQNMDSKITEKKDQDVKELLDSDSK 870

DB 67 DKRCRSRPPKIKIEPDSEKDEKVGSDAAGADQNMDSKITEKKDQDVKELLDSDSK 126

QY 871 PKCEPMEVDDMTKESHVNCQESSQVD--VVNVSEGFHLRTSYKKTKSKLDGLLERR 928

DB 127 PKCERTN--GSRXKHENRVTKLSGEFSSRCGCXGFSKDXLQKXNIIQTRWTSXKE 184

QY 929 IKQFTLEEKQRLKIEKLGKTKGTSTNSKNLSESPVITKAKEGCQSDSMRQEQSP 987

DB 185 NXYVTGKRTATRNQVGGWGXGKDYKDSNSTNSKNLSESPVITKAKEGCQSDSDETRTEP 243

RESULT 12

ABB62887

ID ABB62887 standard; Protein; 997 AA.

XX AC ABB62887;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 15453.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX OS pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO2001171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PA Venter JC, Adams M, Li PWD, Myers EW;

XX PI WPI; 2001-656860/75.

XX DR

DR N-PSDB; ABL06990.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX PS Disclosure; SEQ ID NO 15453; 2lpp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA

CC sequences (ABLI01840-ABLI16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 997 AA;

Query Match 4.7%; Score 672.5; DB 22; Length 997;

Best Local Similarity 24.7%; Pred. No. 5.3e-26;

Matches 311; Conservative 145; Mismatches 395; Indels 409; Gaps 51;

QY 1394 DMEIETSEKVKVTSSPITSEESNL-----SNPDFID---ENGLPINK 1432

DB 6 NVEIESTEDSIVTCLNSGNAEDVDMPGWRKRKNOKSKSYIGTKDVLDTDKDIPLNK 65

QY 1433 NENVNGESKRTVITEVTMTSIVATESKTVIKVEKDKOTVVSSTENCAKSVTTTTTT 1492

DB 66 -----QNRFPIT-----ARPKREC-----VKYERETFENGNERV----- 97

QY 1493 VTKLSTPSTGGSV-----DIISVKEQSKTVVTTTDTSLTTGGTLVTMTSVKEYSTRDK 1548

DB 98 ----YSTSSPGRVLLNDAAKLYEQA-----VKTEDEK 126

QY 1549 VKLMKFSRPPKRTSGTALPSYRFPVTKSTKSPVLPNDLKKLARKGGIREVPFYFNA 1608

DB 127 STITK--KPSYSR----YPLISNPLTHKKRSLVLPFRFELLKLLGGSSTNGFHHAA 180

QY 1609 KPALDIWYPSRPTFGITWRYRLQTVKSLAGVSLMLRLWASLRDWDMAAKVPVGGST 1668

DB 181 KNN-TIWQYQCSRPLFTCWSYTSNATSLSSALQRLWLSCLRWDMDIAK-PPSTDGK 238

QY 1669 RTSETSEITTEIKRRDVGYPYIRFEYCIRKICPIGVPEPKETPTPQKGLRSAL 1728

DB 239 HQVTTDTTEIVTLELLKLRHSGRYGKTSYLRKRVVIPLEMPKTREV-TSIRSLR---- 293

QY 1729 RPKRPTPKQTPGVIIETWVAEELEWEIAPAFARVEKEKAQAVEQ--AKKRLQEQKPT 1787

DB 294 KKRKRAESPPQTEPQITEEWDDEKLELWEIKFMGEKQEKARLSAVTSRSASRQLE----- 348

QY 1788 VIATSTTSPTSTSTISPAOKVMVAP--IGSVTTGTWKVLTTRKVGSPATVTFQONKNEH 1846

DB 349 --ASGSNGSNVSTNGALGVACRVQLAPKLSBDVK--EKMEQQLXL---QRAVHQQRK--- 398

QY 1847 QTFATVWKQGSNSGVQVQOKVLG----IIPSTGTSGQQFTSFQPTATVTRPNTSG 1902

DB 399 -----LVATGEITRSVTPVKGVIGSRVIVKNPDGTR----- 432

QY 1903 SGGTTSNSQVITGPQIRPGMTVIRTPLOOSTLGRAITRTPVWQVQPGAPQVMTQIIRGP 1962

DB 433 -----II-----QQAVTQVSRGTG 446

QY 1963 VSTAVSAPNTVSTPGQKSLTSATSTNSIOSASQPPRPOQGVKLTMALQTLTQGHGG 2022

DB 447 ANTAATAA---ASPTVG-----GSTSQS-NFSTSTPHKVQ----- 477

QY 2023 NQGLTVTVIOGOGQTTGQLQLIPQGVTVLPG--PGQQLMOAAMPNGTVORFLETPLATAT 2080

DB 478 -----IIRG-----PDGKVSVRGLNPGQQLVQ--MPDGKHL----- 506

Query Match 3.9%; Score 560; DB 22; Length 5560;
Best Local Similarity 19.3%; Pred. No. 3.7e-19;
Matches 520; Conservative 382; Mismatches 948; Indels 838; Gaps 118;

QY	325	DRRRKYWFLNRLIIEEDTENENKWIYYSKVQVLAELICLDKDYWEAELCKILEEM	384
DB	1997	DKQKKEKEIKEDLREKQERONREK-----ELRDKDLREKEM-REKEQR	2041
QY	385	REIHRHMDITE-----DLTNKARGSNKSFALAAANEEIESTRAKKGIDNV	431
DB	2042	EKELHREKQOREREHREKEQSRAMDEQEGRGM-----RELSSYQKSKMDIAGE	2093
QY	432	KSPETEKKNETENDSKAEKNREFEQDSLE--KDSDDKTPDDPPEQCKSE-----	482
DB	2094	AS-----SLTAIDCOHNKENAMDTIAQTPCASPTSPDNTPKERSKLSRNSP	2142
QY	483	VGDFKSEKSNGLSESPGAGKAGSGSTR-----IITRLRNPDSK-----LSQLSQOV	530
DB	2143	VLHKKRLSQQESNHHAGGGCGGSGHQIHEDYVKRIKENSQINISVHSSNQLRNDRR	2202
QY	531	AAAAHEANKLFKEGKEVLYVNSQGEISR-----LSTKK-----	563
DB	2203	DSKEHKSSS-FKEDK-----NSSSHISRPHGCGSSASSKHHRRDKHHKQGSASSET	2256
QY	564	-----EVM-----KGNINNYFKLQ-----EGYRVYINQYSTNSFALNK-----HOHR	603
DB	2257	NSSIEVYVDPISQTKNLNLTSEBELQSHQPKREKEREHFSSHANSSSRHKKSRDRHHHR	2316
QY	604	-----EDHRRRLHAKFCLTPAGEFKWNGSVHSGKVLITLTLITOLE	649
DB	2317	EKKRHSVARSNTNDEETPOOHNPHR-RISAAG-----SGSAG-----ELSSAA	2359
QY	650	NNIPSLFHPNASHRANKIKAVQCMKPREFALALAILCAVKVPMPLPWREFLGHTR	709
DB	2360	TNFSCKLHQ--HRRSVERKSSRGSDGHHSSKSL-----RAK	2398
QY	710	LHRMTSIEEEK-----EKVKKKEKQEBEETMQQATWYKYTF	747
DB	2399	LAMLSSADSDTDASKKHSIFDIPDCPNVMDYKVKARSKCNMQOAEKKIKAFKS-	2457
QY	748	PVKHVVQKQGEYRVYTGWGS--WISKTHYRVFVKPLPGNTNINYSKLSBGTNNMD	804
DB	2458	-----QLKQSAKKRSTSDGSDTEFEDRQH-----RNSGSSSFHGR--YFGLSSDD	2505
QY	805	ENMDESKRCKSPKIKIEPDKSEKDEKVGSDAAKADQNE-----MDISKITERKQDQ	859
DB	2506	DDDEETHORRIS-----SDSDAHEGGQDNQCGASTLADANRYRQMQ-QN	2547
QY	860	VRELLSDSDKPKCEPMEVDDDMKTESHVNCQESSQOVVVNVSEG-----FHLRTSY	912
DB	2548	LRRLCDGDDS---SED--EIRRVNMKSHFGKRNSTRIASDSSESQSPAPDLTIKQEH	2602
QY	913	-----KKTKSKLDGLLERRIKQFTLEEKORLEKIGKIGKIGTSTNSKNLS	964
DB	2603	PIAPAQEIKEQLSDEQFKSHDNSSEIER---KLKTEREI---KTELGDYNS	2654
QY	965	ESPVITKAKGQOSDMRQEQSPNANDQPEDLIQOCQSODSS-----VLRMSDFS---	1015
DB	2655	EYTYTKLKEYSPETRKHKHKKRR-----LKSSSTADTSAATPLVMTPLTSPID	2706
QY	1016	-HTTNKLYPKDRVLDDVSI-----RSPETKCPKQNSIENDIEKVSQDLAS	1059
DB	2707	VHSSECKTKFDFDLKTECSSIPIEISAGERKKERKKERKRLNNTEATVPN---	2763
QY	1060	RGQEPKSKTKGNDFFIDDSKLASADDIGTLCKNKKPLQEESTDIVSSSKSALHSSVP	1119
DB	2764	-----SPTNDT-----SSEKUSKEERHKKKSKSKSMDNSCNTKIYNSGA-H---	2808
QY	1120	KSTNDRDATPLSR-----AMDPEKLGCDSESNSTLSENSDTSVIOQDSSEDM	1167
DB	2809	STSPSLPAFTSAPTSAQTSKRGEDKMEFIFGLIISDEESQFPPEQATNKKDIIPSS---	2864

QY	1168	IVQNSNESISEQFRTRREQDVEVLEPLKCELVSGESTQNCEDRLPVKTEANGK-----KPS	1223
DB	2865	-VSTTGPVSAALQTYKQ---Ep-----STPNS-----KNEEAHIQLTVEHEPE	2903
QY	1224	QOKKLEERPYNKSDQIKLNTTDDKKNENRESEK-----KQORTSTQINGKONKPKIYL	1279
DB	2904	QOQOLER---SRUSGGSSSSHADREKRREKREKRSQREQQOIHOKSSK-----	2955
QY	1280	KGECLEIBESRVVGVNPKVNNINKIIPENDIKSLTVKESAIRPPINGDVIEMDFNER	1339
DB	2956	-----VETKVDNDSVDMDEAGRALEAO-----LMSDFDTK	2986
QY	1340	NSSETKSHLLSSSSDAEGNYRDSLETLPSTK---ESDSTQTTTSPASCPESNSVQV---	1392
DB	2987	PISE-----EATPSTAATYRSDMTDFRESDN-EDNNSVDMTKQGV	3036
QY	1393	--EDMETETSEVKKVTSSPITSEESNLSNDFIDENGLPINKNENNVNGESKRKTIVITEV	1450
DB	3027	KSQOQOQHKSKDKKKKKRKEKQEKLL-----QOORRESLPNVA	3067
QY	1451	TMTSTVATESKTVIKVEKGDQKQ---VVSSTENCAKSTV-----TTTTTTV	1493
DB	3068	STSSAPPTPKCLTVNVQAASKHADQLDAKHSSPPVKPSPCLIGDDDDDALHTPK	3127
QY	1494	TKLSTPSTGGSVDIISVKEQSKV---VTTTVDLSLT-----TTGGTLVTS---	1536
DB	3128	AKPTTSSRGNDGLTPSRKXPLIPKPTTIANSTLSTQSAETPVSSGTVISSSALA	3187
QY	1537	-----MTVSKKEYST	1545
DB	3188	TTPTSTTAAGVSAAPGLDNPSTASAOCKKESFIPGFDQGLDDRISAVQSISAEFNS	3247
QY	1546	-----RDKVKLMKFSRPKTRSGTALPSYRKFVTKSTRK-----	1579
DB	3248	TSLLDNIADPKIPVAPPRATKPLDKLEESKSRVTISQEETESAVSALLGESFGTSST	3307
QY	1580	-----SIFVLNDLKKLARK-----GGIREVYFNYNAPAL	1612
DB	3308	DYSLGDMDESSVNELETPTLVIAEPDEEAALAKAETAGETAPASILEEP-----	3363
QY	1613	DIWYPSP-----RPTFGITWYRL-QTVKSLAGVSLMLRLLWASLRWDDMAKVPVPG	1665
DB	3364	EAPDPDPEAEISEPVEVLDPEELNKAQVSLKHEDMM-----DIKADTQSE	3412
QY	1666	GSTRTETSEITTEILIKRRDVGPGYRFEYICIRKICPIGVPEPKETPTQKGLRS	1725
DB	3413	RLQIDT-DREENPDE---ADSGP-----SLKIDETVQSSSSPEKSISNN	3454
QY	1726	SALRPKRPETPKOTGPVIEIETWVAEELELWEIRAFARVEKEKAQAVEQAKRLQOK	1785
DB	3455	S-----PTPRETANIDI-----PNVESQPKLSNESTPQPSVITKL	3489
QY	1786	P-----TVIATSTSTPTSTTSTISPAQKVMVAPISGSV---TTQTKMVLTKVGS	1835
DB	3490	PFLDTKTVPAGLPSPVPKIEPTTISKLOPLVQVQTVLPAPHSTSGSISANSVINLDL	3549
QY	1836	TVTFQONKPHQTFATVWVKOGSNSGVVQVQKVLGIIPSTSTGTSQOTFTFSFQP-RTAT	1894
DB	3550	SNVSSCNTSAASATASASASISFGSPASQANAM---PQASTPKQGPITPQOAIQTSL	3606
QY	1895	TIRPNTSGSGGTTSN---SQVITGPQI---RPG--MTVIRTPLOQSTL---GKAIIRT-	1941
DB	3607	IMQPTTISIEQTPHFPAVPMQVLSPOSHHQQPQGTVMVGRAPSPHPLSPGSGVQASR	3666
QY	1942	-----PVMVQGAPOQVMTQIIRGQPVSTAVSAPNTVSTPGOKSLTSATSTSN	1991
DB	3667	LVQOLSPVGRPMWSQP-SPQO---QVQOQOQHIALITSPSSNISLSPASTTRVLSSNS	3722
QY	1992	QSSA---SQPPRPQGGQVKLTMAQLTQLTQGHGNGOGLTVVIGQGGQGTGOLQILPOGV	2048
DB	3723	PTTSKVNYSYQPRNOQ-----VPQOPSPKSAEVAQVTPQMLMT	3758
QY	2049	VLPGPGQOLMOAAMPNGTQVORFLTPLATTTATTTTSTTTTAACTGQORQSKLSPQM	2108

3759	I---P Q K T P O V P H P T --- I S K V T V Q P Q A T Q S V A S P P L G S ----- L P P H K 3805
2109	QVHHQDKTLTPPAQSSVGPAPAKAQPT-AQPSARBPQPTQPSPAQPEVQPTQPEVQTOTTVS 2167
3806	NVH----- L N A H Q N ----- Q Q O P V T A K T A H Q H Q H Q O F H Q Q I Q R Q H Q O ----- 3849
2168	SHVPSBAQPTHAQSSKXPVAAQSQPOSNVQGGSPVRVQSPSQTRIRPSTPSQLSPGQ-QS 2226
3850	----- Q Q U H G Q S O --- Q T S A P Q H Q H Q Q A Q Q Q H N Q O H L Q L H A Q H P T Q K H 3901
2227	QVOTTTTSQPTPIQPHSTSLQIPSQGQSPQPOVQS----- S T Q T L S S G G T L N Q V S V S P R Q 2283
3902	QAOQFNQ-- Q I Q H Q S Q --- Q H V Q Q N Q A Q Q H L S Q Q H Q S Q Q L N Q --- Q H Q A Q 3952
2284	LQIQQPOP-QVIAVPPQLQOO----- V V L S Q I Q S V A Q I Q A Q S G V P Q I K 2329
3953	QQLQIQIKLQOMHGPQQQKSPQGVGHGGSTSFASQQHNSQLPAR----- G V P Q Q --- 4004
2330	LQLPIQIQSSAVQTHQIQNV-VTVQAASVQEOLORV-- Q Q L R D Q Q K K K Q Q O I E K R H 2386
4005	-QH P Q L S H S C K P N T L S V N G V Q P P A I --- L T R V G S H S Q N Q Q L P H Q S S G H P 4060
2387	TLQASNOSETIQKQ----- V V N K H N A V --- E H L Q K K S M T P A E E R N O I T C N Q V K Y I L 2440
4061	QKQLSSPGANLPLQTLPLNVTQNTPKIIVQOHIHVAQNVPPQPTQGNAIHYPQNO-QKDST 4119
2441	DKIDKEKQAARKKREESVEQRKSNQATKLSAL----- L F K H E Q L R A E I L K R A L 2493
4120	PPGHVEPTPMSAQKTSSESYSVIRTPPTTGLAVISANTVGSLLTEENLIKISQKODEL 4179
2494	LKDQLQTEVQ----- E E L R D K I K K ----- E K D L Q L A Q A T A V A A P P V T P V L P A P 2541
4180	IEQSK-EVUSDTSWAKEVNDISVIKKLDTPLASKDAKRAVEMQAI----- 4224
2542	PAPPPPPPPGVOHTGLLSTPTLPVASQKRRKEEKDSSSKSKKKM 2589
4225	-A P A I N P O G N O --- S M A O E T A L P T S M S V N N S N D H D T E T R O L 4269

Search completed: September 24, 2003, 01:11:28
Job time : 64.7816 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	3363	23.5	810	2	G01252	small GTP binding
2	871.5	6.1	1711	2	T21432	hypothetical prote
3	522	3.6	2447	2	T16870	hypothetical prote
4	494.5	3.5	5327	2	T13564	microtubule-associ
5	459	3.2	5170	2	T15348	hypothetical prote
6	427.5	3.0	3759	2	A35085	trithorax protein
7	419.5	2.9	2232	2	T34434	hypothetical prote
8	413.5	2.9	3507	2	T34513	hypothetical prote
9	412	2.9	2897	2	B48666	cell proliferation
10	404.5	2.8	452	2	T21435	hypothetical prote
11	404.5	2.8	3256	2	A48666	cell proliferation
12	404.5	2.8	3488	2	T34418	hypothetical prote
13	402	2.8	1630	2	A35377	ascites sialoglyco
14	401.5	2.8	1367	1	A84878	glucan 1,4-alpha-g
15	399.5	2.8	405	2	T21433	hypothetical prote
16	399.5	2.8	510	2	D14330	hypothetical prote
17	399.5	2.8	2441	2	T71623	erythrocyte membra
18	397	2.8	2346	2	T13829	TPR homolog - frui
19	395.5	2.8	2271	2	F90073	hypothetical prote
20	394	2.7	2938	2	T30249	cell proliferation
21	387.5	2.7	1229	2	T25697	hypothetical prote
22	382.5	2.7	1589	2	T13606	hypothetical prote
23	382	2.7	2453	2	S60254	nuclear receptor c
24	375.5	2.6	3924	2	S37431	ankyrin 2, neurona
25	374.5	2.6	2116	2	A26655	myosin heavy chain
26	374	2.6	4377	2	A55575	ankyrin 3, long sp
27	373.5	2.6	2541	2	T29340	hypothetical prote
28	372	2.6	5105	2	T32650	hypothetical prote
29	367.5	2.6	3259	1	A56539	giantin - human

481 SE-----482
481 SEPTVEYDGNVSYANLGDNTTATSETPSEGRSPVGCSETPDSSNNAEKKVASEL 540
483 -----482
541 QDVPEPNTKCESNTSATTTSTQPNLENSNSELSSQESAKAADPENGERSHT 600
483 -----VGFKSEKNGELSESPGAGKASGSTRITRLRNPDSKLSQKSOQVAAAA 534
601 PVSTOEIVGDFTEKSTGELSESPGAGKASGSTRITRLRNPDSKLSQKSOQVAAAA 660
535 HEANKLKEGKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLGOEGKYRVYHNOYSTNS 594
661 HEANKLKEGKEVLVNSQGEISRLSTKKEVIMKGNINNYFKLGOEGKYRVYHNOYSTNS 720
595 FALNKHQHREDBDKRRHLAHLKFCITPAGEFKWNSVHGSKVLTITLRLTITOLENNIPS 654
721 FALNKHQHREDBDKRRHLAHLKFCITPAGEFKWNSVHGSKVLTITLRLTITOLENNIPS 776
655 SFLHPNW 661
777 TSLHPSF 783

RESULT 2
T21432
hypothetical protein F26H11.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21432
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21432
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1711 <WIL>
A:Cross-references: EMBL:Z81515; PIDN: CAB04197.1; GSPDB: GN00020; CESP: F26H11.2
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.2
A:Map position: 2
A:Introns: 56/3; 154/2; 373/2; 417/3; 505/2; 596/1; 703/2; 847/3; 922/3; 1034/2; 1078/2;

Query Match 6.1%; Score 871.5; DB 2; Length 1711;
Best Local Similarity 17.7%; Pred No. 5, 3e-25;
Matches 420; Conservative 299; Mismatches 618; Indels 1041; Gaps 71;

4 EEEEE---DGAETQDSEDE---EDEMEEDDDSDYPEEMEDD---DDDASYCTE 52
169 KRQEDLIYMDSESEEESSDDEFLNEDQVQVEEELNLTIDIKIEGLDEENKYC-- 226
53 SFSRSHSYSTPGRRKPRVHRPRPILAE---KDTPPLEFPKSSDLAVPNEHIMNVAI 110
227 -----PWDEDPASLPKLELPSSODIPPTASIMDAVEI 261
111 YEVLNFGTVLRSPREFDCAALVSEQCTTMAEMHVVLKAVLREEDTSNTTFFPAD 170
262 YELRSYHRTLRITPTTFEDFCAALISHNNSCIMAEVHALLNRLNCLKSDDEBQTHSYVTE 321
171 LKDSVNSTLYFDIGMTWPEVLRYVYCESKEYHVLVYQEA-----EDPYGPVENKI 222
322 TNSNVNIMIHMDTLTYABIARQYEA-----YPFADASYRDAINVDNVPFVGYDANI 374
223 K-----VLQFLVDQFLTNTNAREELMSGVIOYDDHCRVCHK-LGDLIC 265
375 QRFSEFFNKKHVRVLVLLFSYRFLYSSEFKLVNNGKFNQDENCRCVCGKSGRVVG 434
266 CETCSAVYHLECVPKPLEEVEDEMOCEVCAH-K-VPGVTDCAVETQKNKPIRIHEPTGY 324
435 CTQCEAAFHVEC--SHLKPFPE-VLCNICKKNSAVRGVLPDPAEAVDREP--LRSQPIGR 489

QY 325 DRSRRKYWFLNRLIIIEEDTENENKKIWIYSTKVQLAEIIDLCDKDYWEAECLKILEM 384
DB 490 DRYGRYWFIVRLVQVSLDETE---LYYSTVPQYQLQKLDRTYYEKLCDTIRL 545
QY 385 REIHMHMDITDLTNKAGSNKSFLLAANBELISIRAKK---GDIDNVKSPE-ETEKD 440
DB 546 IDEFLQMALTVEMTSERR-----EAALETWVKRQILGYDFAEATTPQYLYLHD 594
QY 441 K-NETENDSKDAEKNEEFEDQSLKSDDKTP-----DDDEQCKSGVGVDFKSEKNSG 493
DB 595 SMKRMAIIIRDCAQKQVQKVELEPVGQSPVKQVQVEDSILPESMIGIF-----647
QY 494 ELSPEGAGKAGSGSTRITIRLNPDSKLSQKSOQVAAAAHEANKLFEGKEVLVNSQ 553
DB 648 -----DAKL-----INTFWSGG-----659
QY 554 GETSRLSTKKEVI-----MKGNTN---NYFKLGQSGK---YRVYHNOYSTNSFALNKHQ 601
DB 660 -----ATOEELEYEQFVDISDNFDAPSANLRWMDGENDQTFWTYTYNYSRNEMSESFLT 713
QY 602 HREDHDKRRHLAHLKFCITPAGEFKW---NGSVHGSKVLTITLRLTITOLENNIPSSFL 657
DB 714 RKKAADKKYMASKF--AQIDNFDWVAKNRQFYGDASLHCKFIMWTLOQVIKNIPIDLM 771
QY 658 HPNASHRANWIKAVQMSKPRFALALALECAVPVVMPLPIWRFLGHTLRLHRTSIE 717
DB 772 HRKWEFAKGFDELSVADDDYKKLVCTLLKLDCAVAKTTFMPQWNGGLQGLRERITVQ 831
QY 718 REEKEKVKKEKK-----QEEETMOQATWVKYTFPVKHQVWKQGEYRVYTG---YG 767
DB 832 RENFMKQQKKKIDADALTKDLDSDSFVRVYVNMKPKWPNTY--LRQGETYRNAGKSGM 890
QY 768 GWSWISKTHYRVFVKPLPGNTNRYNRSKLEGTNNMDENMDSDSKKCSRPKKIKLEPD 827
DB 891 GWAWVAKYVEKWI-----QVPESPKPLAVTVEIKEITSY 926
QY 828 SEKDEVGSDAAKAGDAQENMDISKITEKDDVYKELLDSDSDRCKPEEPMEVDDDMKTES 887
DB 927 SNRK-----ARLELLVSKIT-----942
QY 888 HVCQESSQDVVNVSEGFHLRTSYKKKTKSKLDGLLERRIKOFTLEEKORLEKIKLEG 947
DB 943 -----KKRQSG-----949
QY 948 GIKGIGTSTNSKNLSESPVITKAKGCCSDSMRQEQSPNANNDQPEDLIQCSQSDSS 1007
DB 950 -----GKSSKKPTFELT-----NGCYSPPSCRS-----971
QY 1008 VLRMSDPHTTNKLYPKDRVLDVSIKSPETKCPKQNSIENDIEEKVDSLARGQEPKTS 1067
DB 972 -----NPNRKC-----YSP---980
QY 1068 KTKGNDFIDDSKLASADDITGLTICKNKKPLIQEESDTIVSSSKSALHSSVPKSTNDRDA 1127
DB 981 -----MCRN-----984
QY 1128 TPLSRAMDFEGLCGDSESNSTLENSDTSVIQDSSDEDMIVQNSNESISEOFRTREQDV 1187
DB 985 -----984
QY 1188 EVLEPLKCELVSGESTGNCEDRLPVKGTANGKKPSQOKKLEERPVKNCSDQIKLKNTWD 1247
DB 995 -----GYLVSQAKQAHDERKLEE-----1001
QY 1248 KKNENRESEKKGORTSTFQINGKDNKPKYILKGECLKEISESRVSGNVEPKVNNINKI 1307
DB 1002 -----1001
QY 1308 IPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETSKSHLLSSSDAEGNYRDSLETLP 1367
DB 1002 -----1001

QY 1368 TKESDSTQTTTPSASCPSNSVNOVEDMEIETSEVKKVKTSSPITSEESNLSNDFIDENG 1427
Db 1002 -----
QY 1428 LPINKNENNGESKRKTIVTEVTMTSTVATESKTIVKVEGDKQT VVSSSTENCAKSVT 1487
Db 1002 -----SGLGEEK-----
QY 1488 TTTTTVTKLSTPSTGGSDIISVKEQSKTVVTTVTDLSLTTTGGTLVTSMTVSKEYSTRD 1547
Db 1010 -----
QY 1548 KVKLMKSPRKTRSGTALPSYKRVTKSKTSIFVLPNDLKLARKGGIGREV--PVFN 1605
Db 1010 -----AMPIDQTFSTKRGKSKTSIFVLPNDLKLARKGGIGREV--PVFN 1605
QY 1606 YNAKPAIDWIPSPRPTFGITWYRLQTVKSLAGVSLMLRLWASLWDDMAAKVPPGG 1665
Db 1055 AGIKSNLLIWPYAPRPTLDLCKWQTLNARSLHVALQKLIWSSIKFNEFD---PDDT 1111
QY 1666 GSTRTSET--EITTEIIRKRDVGPYIRFYCIKRIICPI-GVPEPKETPTPQKRG 1722
Db 1112 HPDRRVVDTSPHDERRIIRKEMPPYQYERYEMEIEIIPLYDEPEEDESNLRRNG 1171
QY 1723 -----LRSSALRKRPET-----PKQT-----GPVITWVAEELE 1754
Db 1172 GSSEFSRSARKRKQRHEFLSLKFGNYPKKNAFRSLDNRRAIRREWDGVTLK 1231
QY 1755 LWE-----IRAFARVEKEKAQV-----EQAKKRLQOKPTVIATSTTSPTSST 1800
Db 1232 VFEIKDWYKWRARAEAKTAKKLEATRAKAQAKADEERRRQQOQOQSVARIPV-PMHS- 1289
QY 1801 TSTISPAQKVMVAPISGV---TTGTKWLTITKVGSPATVTFQON-----KNFHQTF 1849
Db 1290 ---LIPSENNV-PYLGSOQOQRRPNERGFLKYNSSSVPOAHGVSATPPPGYHQ 1345
QY 1850 ATWYKQOSNGVVOQKVLGIIPSTSTGTSTQFTSFQPTATVTRPNTSGSGGTTSN 1909
Db 1346 PNIIRQAGYNQ-----LPRKPTTSPFNQFQ-PP-VATIPTTPQLRAAGADG- 1390
QY 1910 SOVITGPIRPG---MTVIRTPLOQSTLGAIRTPVMVOPGAPQVMTQIIRGQ--- 1961
Db 1391 --VVRVNMVTCNKSTVNTSTPYQ-----ALNRQYQLOQOQOQFAVRLTNGYHFMD 1443
QY 1962 -----PVSTAVSAPNTSVSTPQOKSLTSAT----- 1986
Db 1444 GTRGGGRNPSVMQHRQLPQNRALQRPFGESTTEMRRTVEAIPDNDGDDEQPPVPIRYD 1503
QY 1987 STSNIOSSASQPPRQOQOVKLTWALQTLTQGHGGNGLTVVTOGQOTTGQLQL----- 2042
Db 1504 PTSNFDQAQOQOHPQSRPVYSTPAQMIITTPQGVKX--NVILMKASDGTQKRWLKPQ 1561
QY 2043 IPOGVTVLPQGOQLOMAAMPNGTVQRFLEPLATTATTTTSTTAAAGTEOROS 2102
Db 1562 FPGGTVI--STGQVRYRQPTAVQOQLYT-----ATPCTRVVRIPNANGAPRQODH 1613
QY 2103 KL-----SPQGVHODK--TLPPAQ-----SSSYGPAKAQOQTAQPSARP 2140
Db 1614 QVMRRVYQASPRAMEYMDQGTTPPGQOQRYVLOGGNSGTPPNPPKVSRRGGRGGLT 1673
QY 2141 QPQOQPSQPAQVQTOPEVQT---QTTVSHVSEBAQ 2175
Db 1674 MQMVQOQOQHNPAPHYDPPDDATGFAVSTTTEQVPEQ 1711

RESULT 3

T16870

hypoetical protein T13H2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000

C:Accession: T16870

R:Wu, X

Submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T13H2.

A:Reference number: Z18593

A:Accession: T16870

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2447 <WUX>

A:Cross-references: EMBL:U39653; NID:gl049397; PID:gl049400; PIDN:AAB52494.1; GSPDB:G

A:Experimental source: strain Bristol N2; clone T13H2

C:Genetics:

A:Gene: CESP:T13H2.3

A:Map position: X

A:Introns: 18/3; 135/3; 176/3; 496/1; 693/1; 816/1; 1053/1; 1131/3; 1181/3; 1308/3; 1

C:Superfamily: RING finger homology

F:158-207/Domain: RING finger homology <RRN>

Query Match 3.6%; Score 522; DB 2; Length 2447;

Best Local Similarity 18.5%; Pred. No. 9e-12;

Matches 502; Conservative 413; Mismatches 1033; Indels 760; Gaps 114;

QY 9 EDGDAEFTQDSEDEDEDEDEDDSDYPEEMEDDDDDASYCTESS----- 54

Db 2 DSDPGPSTKSARKDAENAEENTSDSSDSEVSASEKSESRSPSEKKKVVITRIVPVRP 61

QY 55 -----FRSHSTYSSTPGRRKPRVHRPRSPILEEKDIPLEFPKSSDLMPNHEIMVVI 108

Db 62 PTRDKGHRVNLLESGNESETKSLYQRAKEGIPSYKGPKEIKLPTTSEQYIDLEEYLMN-- 119

QY 109 AIYEVLRNFGTVLRSLSPREFDECAALVSQOCTLMAEMHVLLKAVLREEDTSNTTGP 168

Db 120 -----PARMEG-----RELTLNA----- 132

QY 169 ADLKDSVNSTLYFDIGMTWPEVLRVYCESDREYHHVLPYQAEADYYPGVENKIKVLQFL 228

Db 133 --YDAVRNKYNLPG-----KSVCEAD-----LQKV 156

QY 229 VDQLFTTNIAREELMSEGVIOYDDH--CRVCHKGLDCLLCCTCSAVYHLECVKPLEEVP 286

Db 157 IGSE-SCDVCQELIQGSIMTKCGHRFCDCILVAFMRSGNTCTPCTRONLGSKRELOQDP 215

QY 287 E-DEWQCEVCAHVKPGVTDCVAETQKN-----KPYIRHEPIGYDRSRKYWFLENRR 337

Db 216 RFDQLIYQVVSRSIVG--RMAENREHEKDVYFGKGYIEG--GSDWNRKYGDIPNSK 270

QY 338 LIIEEDTENNEKKI-WYYSTKVQLAELIDCLDWDYEAELCKLILEEMREEIHRHMDITE 396

Db 271 LKAPRLKSGAKKIRWFES-----DED---GSRKVMESKGA-----PKE 310

QY 397 DLTKARGSNKSLAANEEILESTRAKKGIDID---NVKSPETEKKNETENDSKDAEK 453

Db 311 DDTNYLENDKEGTSVAEKEVLE-----EGEMDFPIEKSSDEEQTDLDEEESMLDSDF 365

QY 454 NREEFQDQSL-----EKDSDDKTDDDPQEGKSEVGVDFKSEKSNSELS 498

Db 366 EISDNEDVSKSCSTSKTTNRSDSSSDSDNRDNLQKKRKNKRVKPTDSDSVN 425

QY 499 PGACKGASG---STRIITRLR-----NPDSKLS-QLKSQQVAAAAHAANKLFKEGKVLV 549

Db 426 ESFDEADSGEVVATKLKESKKKCGRPKKFAPELIEGDIPTPSDLSITSSDERDDNA 485

QY 550 VNS-----QGEISRLSTKKEVIMKGNINNY-----FKLQGEKRYVYHNOYSTN 593

Db 486 ADYAFVQKEFNRDPRRDGHGHEKDKLYNFDIMDNHVDKRFKDEGIEHVISD--DSN 543

QY 594 SFALNKHQHRDHDKRRHLAKFKCLTPAGEFKWNGSVHGVKVLITLTLTLTQLENNIP 653

Db 544 S-----EHESDEAEDRESI-----DSEHEKEI----- 566

QY 654 SSFLHPNASHRANWIKAVQCMKSPREFALALALECAVKKVPMPLPIWREFL----GHTR 709

Db 567 SKFL-----SHR-----QPLNPNTS-----VDDCCQVITVVKDKVQSAITSPGTS 609

QY 710 LHRMTSTEREKEKVKKKEQEBEETMQQATWTKYITFPVKHQWVKQGEERYVTGCGW 769

A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1

C;Genetics:

A;Cross-references: FlyBase:FBgn0025392

A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1

A;Note: EG:4984.1

C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.5%; Score 494.5; DB 2; Length 5327;
Best Local Similarity 17.7%; Pred. No. 2.6e+10;
Matches 527; Conservative 478; Mismatches 1189; Indels 791; Gaps 119;

QY 1 MVSEEEEDGDAETQD--SEDDDEDEEEDDDSDYDPE-----MEDDDDDASYC 50
DB 1751 LASEASRPASVAESVYDDEAKESKEESRRESVAEKSPLPSKEASRPASVAESIKDAEKS 1810
QY 51 TESSFRSHSYSTPGRRPRVHRPRSPILEERDIPPLEFPKPSSEDL-----VPN 101
DB 1811 KEES-RRESVAEKSPLPSK-EASRPASVAESIKD---EAKSKEESRRESVAEKSPLPS 1864
QY 102 EHTMVIATIEVLRNFGTVLRSPFRFEDFCALVSOEQCTLMAEMHVLLKAVLREEDT 161
DB 1865 KEASRPASVAESIKD-----EAKSKEE-----SRRESVAEKSPLPSKEA 1904
QY 162 SNTTFGPADLKDSVNSTLYFIDGMTWPEVLRVYCESDKYHHVLPYQAEADYPYGPVENK 221
DB 1905 SR----PASVAESIKDE-----EAKSKEESRRESVAEKSPLPSKEASRP 1944
QY 222 IKVLQFLVQFLTNTIARBELMSEGVIQYDDHCRVCHKLGLDLCCTCSAVYHLECVKPP 281
DB 1945 ASVAESIKDE---EAKSKEESRRESVAE-----KSP 1972
QY 282 LEEVPDENOCEVCVAHKVPGVTDCAEIOKNNPYIRHEPIGYDRSRRKYFWLNRLLIIE 341
DB 1973 L---PSKEASRPASVAESIKD-----EAKSKEESRRESVAE-EKSPLPSKEASRPASVA 2022
QY 342 EDTENENEKIMWYSYFKVOLAEILDCLDKYWEAELCKLLEEMREIHRHMDITEDLTNK 401
DB 2023 ESIKDAEKS-KEESRRESVAEKSPLPSKE---ASRPASVAESIKDAEK-----SKEESRR 2075
QY 402 ARGNSKSFLL---AANEEILIESIRAKKGDIDNVKSPETEKKDNETENDS-KDAEKNRE 456
DB 2076 ESAAEKSPLPSKEASRPASVAESVK-----DEADKSKEESRRESMAESGKAQS 2123
QY 457 EFEDQS-----LEKSDODTTPDDDPQKGSEVGDFFKSEKNGELSPGAGKA 505
DB 2124 IKGDQPLKEVRPESVAESVKKDPVKSKPEPSRRESVAGSVTADSARD--QSPLESKGA 2181
QY 506 SGSTRIITRLRNPDSKLSQLKSOQVAAAHAANKLFKEGKVL-----V 549
DB 2182 SRPESVYDSVKDEAEKQESRRESKTESVIPPRAKDDKSPKVLQPVSWMTETIREDADQPM 2241
QY 550 VNSQGEISRLS-----TKKEVIMKGNINNYFKLQGEQKYRVYHNOYST 592
DB 2242 KPSQAEERRESIAESTIKASPRDEKSPLASKEASRPGSVAESIK-----YDL 2288
QY 593 NEFALNKHQHRDHDKRRHLAHKFCFLTAPAGEPKWNGSVHGSKVLITSTLRLLITOLENNI 652
DB 2289 DKPQIHKDKSTEHRSRESLEDKSAVTS-----EKSV--SRPLSVASDHEAAVATEDDA 2340
QY 653 PSSF-----LHPNW-----ASHRANNIKAVQMS-----676
DB 2341 KSIISPKDKSRPGFVAETVSSPIEATMEFSKIEVVEKSSLSLALSGGGGKLQTDSSPV 2400
QY 677 --KPREFALALALECAVPVVMPLPIWREFLGHTR-----LHRMTSI 716
DB 2401 DVAEGDFSHAVASVS-TVPTLTTPAELAQIAGAAKTVPSSPLDEALRTPAPAHISRADSP 2459
QY 717 EREEKEVKKKKQEEETMQOATWVYKTFPVKQHVWKQKGEYRVTVGYGWSWISKTH 776
DB 2460 ABCASBEITASQDKSPQVLKESRPAPWABSKDDAQL-KSSVEDLR-----SPVASTE 2511
QY 777 VYRFVPKLPDNT-----NWNYSKLEGTKNMD-----804

DB 2512 ISR--PASAGETASSPIEEAPKDFAEFEQAEKAVLPLTIELKGNLPTLSSPVDVAHASVQ 2569
QY 805 -ENMDESDRKCRSRS-----PKKIKIEPDSEKDE-----VKGSDAK-----840
DB 2570 PAELSKVDTIEKTASSPIDEAPKSLIGSPAEREPEPAESAKDAEAEVSEKSKDASRPSPSV 2629
QY 841 ---GADONEMDIS---KITEKKDDQVKELLOS-----DSDKPCKKEPMEVDD 881
DB 2630 ESTKADSTKGDISPSPESVLEGPKDDVEKSKESRPPSVSASITGDSKTQV-SRPASVVE 2688
QY 882 DMKTESHVANCOESSOVVNVNBSGFHLRTSYKKKTKSSKLDGLLERRIKOFTLEEKORLE 941
DB 2689 SVKDEH--DKAESRRESIAKVES--VIDEAGKSDSKSSSQD-----SOKDEKSTLASKEASR 2741
QY 942 KIKLEGIGIKGIGTSTNSSKNL-----SESPVITK--AKEGQSDSMRQECPNA 989
DB 2742 RESVVESSKDDAEKSRSRPESVITASGEPVPRESKSPLDKDTSRPGSVSVESVTAEDKSE 2801
QY 990 NNDQPEDL---IOGCSOSDSSVLRMSPDSSHHTTNKLYPKDRVLDVVSIRSPTCKPKONSI 1046
DB 2802 QOSRRESVAESVKADTKDKGKSQEASRPSVDELL--KDDDEKQESRRQSITGSHKAMST 2859
QY 1047 ENDIEEKVSDLASRGOEPTK-----SKTKGNDFFIDDSKLASADDIGTLCKNKK 1096
DB 2860 MGD--ESPMDKADKSKPEPSVAESIKHENTKDEESPLGSRDRSVAESIKSDITKGEK 2917
QY 1097 -PLIOQE---SDTIVSSSK-----SALHSVPKSTNDRDATPLSRAMDF-EKGKGC 1142
DB 2918 SPLPSKEVSRPESVGSIKDEKAESRRESVAESVKPESSKDATSAPPSKEHSRPSVLGS 2977
QY 1143 DEESNTLNSDSTVSIQDSSSEED---MIVQNSNESISQFRTREODVEVLEPLKGCELVS 1199
DB 2978 LKDEGD--KTTSRVSVADSIKDEKSLVSOEASRPESE-----AESLKDAAP 3024
QY 1200 GESTGNCEDRLPVKGTTEANGKKPSQOKKLEERPVKNCSDOIKLKNTTDKNNENRESEKK 1259
DB 3025 SOETSRPES---VTESVKDGKSPVASKE--ASRPASVAEN---AKSDADESKEORPELPO 3077
QY 1260 GORTSTFOINGKDNKPKIYKLGCLKEISESRVSVGNVPEPKVNNIKIIPENDIKSLITVK 1319
DB 3078 SRAGSI-----KDEKSPASKDAEAKSKEESRES-----VAEQFP-----LVS 3116
QY 1320 ESAIRPFINGDVIMEDFNERNSETKSHLLSS-----SDAEGNYRDSLETLPETKESDST 1374
DB 3117 KEYSRPASVAESVKD---EAEKSKEESPLMSKEASRPASVAGSVKDEAEKSKEESRESV 3173
QY 1375 QITTPSACSPESNVQVEDMEIETSEVKK-----VTSSPITSEESNLSNDFIDENG 1427
DB 3174 AEKSPULPSKEASRPASVAESVDEADKESRRESGAEKSPLASKEASR-----3223
QY 1428 LPINKNENYNGESK-----RKTIVITEVTMTSTVATESKTVIKVEKGDKQTVVSSSTENC 1481
DB 3224 -PASVAESIKDEAEKSKEESRRESVAEKSPLPSKEASRPTSVAKSVKDAE---KSKES 3279
QY 1482 AKSTVTTTNTTIVTK-LSTPSTGGSDIISVKEQSKTVVTTTVDLSITTTGGTLVTSMTVS 1540
DB 3280 SRDSVAEKSPLASKEASRPASVAE---SVQDEAEK-----S 3312
QY 1541 KEVSTPDKV-KLMKFSRPKKTSGTALPSYRKFEVTKSTKKSIFVLPNDLKLARKGGIR 1599
DB 3313 KEESRRESVAEKSPLAYKEASRPASVAESIKDAEKSKEES-----RRRESVAE 3360
QY 1600 EYPIFYNNAKPALDINWYPSPRPTFGITWYRIQTVKSLAGVSLMLRLLWASLRWDDMAA 1659
DB 3361 KSPASKEAS-----RPT-SVAESVKDEAEKSKEESR-----DSVAE 3397
QY 1660 KVPFGGGTTRTETSETETITTEILIKRRDVG---PYGIRFECIRKTIICPIGVPTPKETP 1716
DB 3398 KSPASKEASRPASVAESVQDEAEKSKEESRRESVAEKSPLASKEASRPASVAESVKDDA 3457
QY 1717 TPQRKLRSSALRPKRPETPKQTG-PVITETWAELELELWEIRAEARVEKEKAQAVEQ 1775
DB 3458 EKSKEESRRESVAEKSPLASKEASRPASV-----AESV-KDEAEKSKE 3499

A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 2.98; Score 419.5; DB 2; Length 2232;
Best Local Similarity 21.88; Pred. No. 5.3e-08;
Matches 339; Conservative 172; Mismatches 633; Indels 413; Gaps 63;

QY 945 LEGGIKIGTGTSSKNSLSESPVITKAKGECQSDSMRQEQ-----SPNANDQPDLLQGC 1001
DB 731 LESTSG-ATYSSGAGTTWSP-----SQSSVSGSGSTSPAASTSGEMTSQGS 781
QY 1002 SQDSSVLRMSDPGHTTNKLYPKDRVLDVYSIRSPETKCPKQNSIENDIEKVSDLASRG 1061
DB 782 TOTPGSSVSTSAALITSTQ-----QSVSTNSPGSTVTRPST-----VSGSTSSG 825
QY 1062 QEPKSKITKGNDFIDDSKLASADDIGTLCKNKKPLIOESDTIV-----SSSKSALHS 1116
DB 826 STVTGSGTEAS---TSGSSVASSSPAPS-TSQNPSTSGSSMITQSPYSPSTSPVES 881
QY 1117 SVPKSTNDRDATPLSRAMDFEGKLGCDSESNSTLENS---SDTVSIQDSSEEDMIVONS 1173
DB 882 STTSPGSGPTTLTST-----PSPQSTIGTQSTSPGISTTSEE-MTSGST 931
QY 1174 ESISEQFTRQDQVEVLEPLKCELVGSEGNCEDLPLVKGTANGKKPQQKKLERPV 1233
DB 932 QTPGSGSTVTPQSTVSD-----STSSGS---TVVSGSTEGS-SSPIPSTQNTNPS 979
QY 1234 NKCSQJIKLNTTOKKNENRESEKKQRTSTFOINGKDNKPKIYLKGECLKEISESRV 1293
DB 980 TSGSSMSTQPPQSSQSTSPVESSTSGATSSS-----GSP-----GTLTSTSPSP 1027
QY 1294 S-----GNVEPKVNNIKIIP--NDIKSLTVKESAIRPFINGVIMEDFNERNSET 1344
DB 1028 SSTIGSSQGSTSPVVISQGSTETPGSTGTVTKPST---VSGSASSGSTATMGSTEA 1083
QY 1345 KSHLLSSDABGNRDSLETLPSTK---ESDSTQTTTPSACPSNSVNOVEDMEITSE 1401
DB 1084 SS---TSGGSSTPNPQSTSPSTSGATSPGSSGTLTSTSPSPQSSSTIGSQGST 1140
QY 1402 VKKVTSPITSEESNLSDNDFIDENGLPINKNENVG-----ESKRKIVIT 1447
DB 1141 VVSTTSGDWTSGGQTQIPG-----STGSTVTPQSTGSGSTSGEITSGGQTQTPRSSLT 1196
QY 1448 EVTTMTST---VATES--KTIVKVGDKQTVVSSSTENCAKSTVTTTTT-----TV 1493
DB 1197 SPAISTSTQOQSVSTNSPGSTVQ-----PSTVIRGSTS--SGSTVTTGSGSTSGSSA 1249
QY 1494 TKLST-----PSTGGSVDIISVKREQSTVTVTTVTDLSL-----TTGCT 1532
DB 1250 TSLSSSSPVPSTQSPNPSTSGSTSTPNPQSTSPVVTGEMTSHGSTQTPSTIGST 1309
QY 1533 LVTSMVTYSKEYSTRDKVKLMKFSRPKKTSGTALPSYKFKVTYKTKSIFVLNPDLLKL 1592
DB 1310 VTQSTVSGSSSGSTVTI---GSEASTSGSSF-----KTPSSISPVPTSS----- 1354
QY 1593 ARKGGIREVFPYFNNAKPALDIWPPRPRTF-GITWYRIQTVKSLAGVSLMLRLWLAS 1651
DB 1355 -----PIPSITTFASSTSGSTISDVSSVSTTSLA----- 1382
QY 1652 LRWDMAKVPGGGSGTSTETSETTEITTEILKRDVGPYGIREFYCIKIKIICIGVPT 1711
DB 1383 -----PLSSSLP-----STVPSSTQSFSTSEGSKASSP-----VPSQ 1417
QY 1712 PKEPTPQKGLRSALRPK-----RPETPKQGPVLIETWAE 1752

DB 1418 TSSTPTNPTGTSSTSSLLSSTISGTOHTTMSKASSGSTSPSTNSQTGTVTMG----- 1471
QY 1753 LELWEIRAFARVEKEKAQAEQAKRLBQKPTVIATSTTSTPTSTTTIIPAQKVMV 1812
DB 1472 -----SSSTSGSVSTSSASTQPMSTSGSSAGSTVASSTAPASSTAPSTG--TMS 1523
QY 1813 APISGSVTTGTTKMWLTTKVGSF---ATVTFQONKNFHQTATVWKQGSNSGVVQVQKV 1869
DB 1524 STSGSTVGSSTISSESTTASASSOTGTVTMGSSST-----SGVSTSSASSTQPM 1573
QY 1870 LGIIPSTGTSQOTFTSFQPTATVTRPNTSGSGGTTNSQVITGPQIRPGMTVIRTPL 1929
DB 1574 STQGSAGSTVASSSTAGLVSTSTV---PSTGTMGSTSG-----TVGSTIS 1618
QY 1930 QOSTLGRKAIIRTPVMVQPGAPQVMTQIIRGQPVSTAVSAPNTVSSPPG-----KSL 1982
DB 1619 ESSTTASASSOTGTVTMGSS-----TSGVSTSSASSTQPMSTSGSSAGSTVASST 1672
QY 1983 TSATSTNSIOSSASPPRPOQGGVKLTMAQLTQTHGNGOGLTVVIOGGQT----- 2036
DB 1673 TGLVSTSTVPSSTGTMGSSSTSGVSTISSESS--TAASASSOTGSTVTMGSSSTSGVST 1730
QY 2037 ---TGLOLQIPQ-----GVTVLPGPGQOLMOAAMP-----NCTVQRFLEPLIATT 2078
DB 1731 SASGQQPMSTSGSSAGSTVSTASPAASSTAPSTGTMTSSTSGTVG-----STMSQS 1786
QY 2079 ATTAATTTTTTSTAAGTGEORSKLSPQMVHQDK-----SSTSSNQMTSGSSVSTAGLVSTVPSSTG 2114
DB 1787 STAASTTSHGTSTVTLGS---SSTSSNQMTSGSSVSTAGLVSTVPSSTG 1842
QY 2115 TLPPAQSSSVG-----PAKAQPTAQ-----PSARPOPTQPSAPQPEVQTQPEVQ 2161
DB 1843 TMGSTSGTGVSTISESTTASASSOTGTVTMGSSSTSGVSTSSASSTQPMSTSGSS 1902
QY 2162 TOTTVSSH-----VPSEAOPTHAQS-----SKPQVAQSQPSNVQCSQSPVRV 2204
DB 1903 AGSTVASTAGLVSTVSTVPSSTGTMGSSSTSGVSTISESTTASSTSSQTG---STVTI 1958
QY 2205 QSPSOTRIRSTPSQLSPGQOSQVQTTSTPIPIQHTSLQIPSGOGQPSQVQST-Q 2263
DB 1959 GSTSGT--NPSSPRSL-----SQI-TITPSPQSTESTQTSLPSSSPSTHVSSEGT 2010
QY 2264 TSSGQTLN-----QVSVSPSRPQLIQOPOQVIAVPLQOQOVLSQ 2308
DB 2011 TMSSGATSGDKMSFLSTGTTFVFSRGSGLATTSAKPSVTCFLMWTQSKIDQ 2067

RESULT 8
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favella, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match 2.98; Score 413.5; DB 2; Length 3507;
Best Local Similarity 19.88; Pred. No. 1.6e-07;
Matches 263; Conservative 176; Mismatches 469; Indels 417; Gaps 48;

QY 1037 ETCKPKQNSIENDIEBKVSDLASRQGPPTKSKTGNDFFIDDDSKLASADDIGTLICKNKK 1096
Db 1941 EKKOPNREKIE--IDENSSSSSQE--KPTTKG----- 1971
QY 1097 PLIQEESDITVSSSKALH--SSVPKSTNDROATPLSRAMDFEGKLGCDSESNSTLENSS 1154
Db 1972 -IVSSTASSTSESTTAPHTVTSISSTTKDMT-----SSKSPENVMTSSES 2018
QY 1155 DTVSIODS-----SEEDMIVONSIESIQFRTREQDVEVLEPLKCBELVSGESTGNCEDRL 1210
Db 2019 PEVSTSSKSTTASSTVSTSPSESSSE-----APLTSSPAT----- 2056
QY 1211 PVKGTANGKPSQOQKLEERPYNKCDQIKLNTDDKNNENRESEKKGQRTSTQING 1270
Db 2057 -----TTEVITESSKSTTPKESSSEITVKLSKS----- 2087
QY 1271 KONKPIYLGCELCIEISRVVGNVEPKV--NNINKIIPENDIKSLVAKESAIRPFIN 1328
Db 2088 -----PEVTESSVKSSPSTPTSTTSQVTSVPTSKSTVLSSEAPVTSTSP 2133
QY 1329 GDVIMEDFERNSSKSHLLSSDAEGNYRDSLETLPSTKESDSQTQT----- 1377
Db 2134 TEV-----HTSSETKPSLSASSTT-----GDTNSTPTSTSSLASVKSTSAPEGTSASVA 2182
QY 1378 -----TPSACSPESNVQVEDMEIETSEVKKVTSSPITSEESNLNDFIDENGLPI 1430
Db 2183 PVKLSLSDPVQSPKTFDATESTVQASSTSSGTSVKSTSSEPSHVTKLSITSS----- 2238
QY 1431 NKNENVNGESKRKTIVITEVTTMTSTVATSKTVIKVEKGDQIVVSTENCASKST---V 1486
Db 2239 NFSSSVPTVPSKTPVPPESTEQPTSTTPSGQSLTPMNSNSEVLTTSSEPHVLSSSLSUDV 2298
QY 1487 TTTTTPVTKLS-----TPSTGGSVDIISVKEQSKTVVTTTVDLSLTGGLTIVTSMTV 1540
Db 2299 SQSSTTPNNLSSTVETPKTSEVSLNSEEPSTTEAPTLSPDILSTTNNLSQSSTV- 2357
QY 1541 KEYSTRDKVKLMKFSRPPKTRSTALPSTRKFVTKSKTSIIFVLPNDLKKIARKGGIRE 1600
Db 2358 ---STEDRSEISENSEKPTSAPELVTSSVTHVASSPD-----VPTSESEPDDLTSGST 2410
QY 1601 -VPYFNYNKAPALDIWYP-----SPRPTFGITWRYRLQTVKSLAGVSLMLRLWASLRW 1654
Db 2411 NIP--EASSKQIISSTPTDITTAEEPKTSMSDPLSTTSNVLSSE----- 2456
QY 1655 DMAAKVPPGGSTRTTSETE---ITTEIHKRDVGPYGIREFYCIKRIICPGVPE 1710
Db 2457 -----STTPESSKSPVSSSTEGISVVTSTFSK-----VPE 2488
QY 1711 TPKETPTPORKGLRSALRPKRPETPKQTPGVIIETWVAEELE-----LWEI 1758
Db 2489 S-----TISSVLEEDLTKT--TPSPILEETTASSETSEPLETDSLAVSVRIHEL 2535
QY 1759 RAFAERVEKEKAQAVEQOAKRLEQOKPVATSTSTPTSTSTISPAQKVMVAPISGS 1818
Db 2536 TTSSENVPE--SESTTTSSESKSPQEPAGILTSVVTPTSSVSLITASE-----IEA 2587
QY 1819 VTGTAKMLTTVKGPATVFOONKFNHOTFATWYKQGSNSGVVOQKVLGIIPSSTG 1878
Db 2588 ITSNTPF-----KQRT-----PITTSKSLVKST-- 2612
QY 1879 TSOQFTSTQPRATVTVIRPNTSGSGGTTNSGOVITGPQIRPGMVTIRPLQOOSTLGKAI 1938
Db 2613 TSPSTVTSSEPSSESTKR-----TTVSTTVST-----TPTTETTTSESL 2651
QY 1939 IPTVMVOPGAPQOVMTOIIRG--OPVSTAVSAPNTVSTPGKSLTSATSTSNIOSSAS 1996
Db 2652 IIT---AAPSKTESTESSEAPTTPAKTSETKPSNVSS-----TSRKSSTENVETSTS 2701
QY 1997 QPPRPQOQGVKLTMAQLTQGHGNGOGLTVVIQOGQGTGLOLIPQGVTVLPQPGQ 2056
Db 2702 Q-----SGSLE----- 2707
QY 2057 LMOAAMPNTVQRFLETPLTATTATTATTTTTV--STTAAGTGEQRQSKLS---POMQVHQ 2112

RESULT 9

B48666

cell proliferation antigen Ki-67, short form - human

C:Species: Homo sapiens (man)

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999

C:Accession: B48666

R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge

J. Cell Biol. 123, 513-522, 1993

A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u

ins.

A:Reference number: A48666; MUID:94043435; PMID:8227122

A:Accession: B48666

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2897 <SCH>

A:Cross-references: EMBU:X65551

C:Superfamily: Kinase interaction domain homology

C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat

F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.9%; Score 412; DB 2; Length 2897;

Best Local Similarity 18.5%; Pred. No. 1.4e-07;

Matches 551; Conservative 391; Mismatches 1062; Indels 980; Gaps 137;

QY 291 QCEVVCVAHVGPVTDCAVEIQKNKPYIRH-----EPIG-----YDR 326

Db .34 ECDIRQLPVSKQHCXKIEHQEAILHNFSTNTQVNSVIDEPVRLKHGVDIIR 93

QY 327 SRKKYFLNRRLLIEDTENENKTIWYSTKVOLAELIDCLDKDYWEAELCKILEMRE 386

Db 94 SFR-----YENESLQNGRKSTFPRKIREQEPARRVSRSSFSSD-----PDEKAQ 138

QY 387 EIHHRMDITEDLTNKAAGSNKSFLLAANEELLESIRAKKGDIDN-----VKSPEETED 440

Db 139 DSKAYSKIITE---GKVSNGPELF---DENLPPNTPLKRGAEPTKRKSLVMHTPPVLKKI 191

QY 441 KNETENDSDAENKRE---EFEDQSL---EKSDDDKTP--DDPEQKSEVGVDFKSEKSN 492

Db 192 IREQPOPSQKQSGSGLIHVEVKAQSLVISPAPSPRKTVPASDORRRSCKTAPASSKSQ 251

QY 493 GELSESPGAGKAGSGSTRIITRLRNPDN-----KLSQKSKQVAAAHAANKLFKEGKE 546

Db 252 TEVPK-----RGGERVATCLQKRVISRSQHDILQMTCSRKRSGASEAN----- 295

QY 547 VLVVNSQGLISRLSTKK---EYIMKG---NINNYFKLGQEGKRV---YHNQYSTN----- 593

Db 296 LIVAKSWADVVKLGAKQTQTKVHKGPQKSMNKRQRRPATPKPVGEVHSQFSTGHANSP 355

QY 594 -SFALNKHQREDHDKRR--HLAHKFLCTPAGEFK-----WNGSVHSGSKVLITSL 641

Db 356 CIIIGKAITEKVVHPARPYRLNPNFISNOKMDFKEDLSGIAEMFTVPKEQPOLT-STC 414

QY 642 RLTIITOLENNIPSSF-----LHPNASHRANWIKAVOMCSK--PREFALALALECA 691

```

Db      415 HIAISSENILGKFOQCTDSEEPLLPTSSEFGNVFFSAQNAKQPSD-----KCS 466
QY      692 VKPVVMLPINREFLGHT-----RLHRMTSIERE-----EK 721
Db      467 ASP-----PLRQOCIRENGNVAKTPRNTKWTSLBTKTSDTETBPSKTVSVNRSRSTEF 522
QY      722 EKVKK--KEKQOEETM-----QOATWVKYTPPVKHQVWKQKQGEYRVTYGGWS 770
Db      523 RNIOQLPVSESEETNIEVICILRGOKATLLQO-----RREGEMKEI----- 566
QY      771 WISKTHYRVFKPLGPNVNNYKSLGKTKNNMDENDESDK-----RKCS--RSPKKIK 823
Db      567 -----ERPFET--YKENIELKEN--DEKMKAMKRSRTWGCKCAPMSDLTDLK 609
QY      824 IEPDSE--KDEVKQ-----SDAKGADONEMOISKITTEKKDQDVKELSDSDKCKE-- 874
Db      610 SLPDTELMKQDARGONLLOTDHAKAPKSEKGTITM-----PCOSL 651
QY      875 EPMVEVDDDMKTESHVNCQESSQDVNVNSE-----GFHLRTS----- 911
Db      652 QP-----EPINTPTHTKQOLKASLKGCVGKEALLAVGKFTRTSGETTHTHREPAGDGKSIR 707
QY      912 -YKKKTK-----SKLGLLERRIKQFTLEBKQRLKLEGGIKLEGKSTNSKNLSE 965
Db      708 TFKESPKQIILDPAARVTGM--KWPRTPKEEAQSLEDL---AGFKELFOTPGPSESMTD 762
QY      966 SPVITKAKGQCSMRQEOSPNANNDQPE--DLTOGCSQSDSVLVRMSDPSHTTNKLYPK 1024
Db      763 EKT---TKIACKSPPPESVDPTSTKQWPKRSLRKADVEEFLALRLKLTSPSAGKAMLTPK 819
QY      1025 DRVLDDVSI-----RSPETKCPKQNSIEN-----DIEKVS 1055
Db      820 PAGGEDKIDKAFMGTPVQKLDLACTLPGSKRQLOTPKEKAQLEDLAGFKELFOTPGHTE 879
QY      1056 DIASRGO-----EPKSKTKGNDFFIDDSKLASADDITGLI--CNKKKPLI 1099
Db      880 ELVAAGKTKTKIPCDSPQSDPVDTPSTKQR-----PKRSIRKADVEGELLACRNLMP-- 931
QY      1100 QUESDTIVSSKKSALHSVPKSTNDRD-----ATPLSRAMDFEGKLAGCDSNSNTLNS 1154
Db      932 -----SAGKAMHTPKPSVGEEDKIIIFVGTVPQKL-----DLTENLT 968
QY      1155 DTVSIQDSSEDMIVQNSNESISIQPRTREQDVEVLEPLKCELVSGSEGTG--NCBDRLPVK 1213
Db      969 GSKRRPQTPKEAQALEDLTGFKELFOTPGHTEEA-----VAAGKTKTKMPCSESSPP 1021
QY      1214 GTEANGKPKSOQKLEERPVKNCSDQIKLKNWTDKKNENRESEKKGORTSTFOINGKDN 1273
Db      1022 ADTPTSTRQPTPLEKRDVQKELSALK-----KLTQTSGETHTHTDKVPGED 1069
QY      1274 KPKIYLKGECLKEISESRVSGNV--EPKVNINKIIPENDIKSLTVKESAIRPFGDVI 1332
Db      1070 KSINAFRETAQKLDPAASVTGSKRHPKTK--ERAQPLEDLAGW--KELFQTP-----V 1119
QY      1333 MEDFNERNSETKSHLLSSDAEGNYRDSLETLPSTKESDSTQTTTSPASCPESNSVNOV 1392
Db      1120 CTDKPTTHEKTKTIACRSQD-----PVDTPTSSKPSQKRSIRK 1158
QY      1393 EDMETETSEVKKVTSS-----PITSEESNLNDIFIDENGLPINK---NENVAGES 1440
Db      1159 VDVEEEFALRRKTPSAGKAMHTPKPAVSGEKNIYA--FM---GTPVOKLDLTENLTGSK 1213
QY      1441 KRKTIVITEVTTMTSTVA-----TESKTVIKVEKGDQKQTVSVSSTENCAKSTVTT 1488
Db      1214 RRLQTPKEAQALEDLAGFKELFOTRGHTESMT---NDKTAKVACKSSQPDLDKNPASS 1270
QY      1489 TTTTIVKLSTSTGSGVDIISVKOSKTVVTTVTDLSLTGCTLVTSMTVSKYESTRDK 1548
Db      1271 KRRLKTSLG--KVGVKELLAVGLTQTSGETHTHTPTGDKGSKAFMESPKQILD 1328
QY      1549 VKLMKFSRPKTRSGTA-----LPSYRK--FVTKSKTSKIFVLPNDL--- 1589

```

```

Db      1329 ASLTGSKRQURTPKGSVEPDEDLAGFIELFQTPSHTKESMTNKTTKVSVRASQPDLDVT 1388
QY      1590 -----KKLARKGIREVPYFNYNNAKPAIDTWPPSPRPTFGITWRYRLQTVKSLAGV 1641
Db      1389 PTKSKQPKSLRKATEE-----EFLAFRKQTPSAG---KAMHTPKPAVGE 1432
QY      1642 SLMLR--LLMASLRWDDMAAKVPPGGSGTRTETSETETETITTTETIIKRDDVGPYGR- 1694
Db      1433 EKDINTFLGTPVQKLDQPGNLP--GSNRRLOTRKEKAQALELT-----GFRELFTQP 1483
QY      1695 -----FEYCIRKIIC--PIGVP--ETPKETPTPQKGLRSS-----ALRKP- 1733
Db      1484 CTDNPTADEKTTIKILCKSPQSDPADPTMTKORPKRSLKADVEEFLAFRLKLTSPAGK 1543
QY      1734 --ETPKGTGPVILLETWAAEELELEWEIRAF---AERVE-----KEKAQA 1772
Db      1544 AMHTPKAA-----VGEK-----DINTFVGPVVEKLDLLGNLPGSKRRPQPKKAKA 1591
QY      1773 VEOAKRRLFOQKPTVIATSTTSPTSTTTISPAOKVMVAPISGVTGTTKMWLTKVKG 1832
Db      1592 LEDLAGFKELFOTPGHTEESMTDDKITEVSKSPQDPVKTP-----TSSKORLKISLG 1645
QY      1833 SPATVTFQKNKHQTFATVWKQGSNSGVVQOKVL--GIIPSTSTGTSTQQTFTSFQPR 1890
Db      1646 K-----VGKVEVLVPVKLTQTSGKTTQT----- 1669
QY      1891 TATVTIRPNTSGSGGT-----TNSQVVI-----TGQIIRPGMTVIRTP-----LOOST 1933
Db      1670 -----HRETAGDKSIAKAFKESAKOMLDPANYGTGMRWP-----RTPKEEAQSLEDLA 1718
QY      1934 LGKAIIRTPVMQPGAPQVMTQIIRGQPVSTAVSAPNTVSSTPGQKSLTSATSTSIQOS 1993
Db      1719 GFKELFOTPDHTEESTTDDKTKTI-----ACKSPPPESMDPTST----- 1758
QY      1994 SASOPPPRQOGVKLT--MAOLTQTOCHGNGOGLTVVIOGOGTTQOLQIP-----QGV 2047
Db      1759 --RRRPTPLGKRDIIVEELSALKULTQ-----THTDKVPCDEDKGI 1798
QY      2048 TVLPGPQOQLM-----QAAMPNGTVQ-----RFLF--TPLATTATTAFTTT 2087
Db      1799 NVFRETAKQLDPAASVTGSKRQPRTPKGAQPLEDLAGLKFELFQTPVCTDKPTTHEKTT 1858
QY      2088 TVSITA-----AGTGE--ORQSKLS--PQMOVHODKTLPPAQSSSVGPAKAQPOTA----- 2134
Db      1859 KIACRSQDPDVGPTTIFKQSKRSLRKADVEESLALRKRTPSVGKAMTDPKPAAGDEK 1918
QY      2135 -----QPSARPQPTQPSQPAQPEVQTPQEVQTPQTVSS-----HVPSEAQ 2175
Db      1919 DMKAFMGTPVQKLDLPGNLPGSKRWQTPKE-----KAQALEDLAGFKELFOTPGTDK 1971
QY      2176 PTHAQSKPQVAAQSQPSNVQSGSPVRVQSPSOTRIRPS-----TP 2217
Db      1972 PT-TDEKTTTKIACKS-PQ-----PDVDPPTASTKORPKRNLKADVEEFLALRKRTP 2022
QY      2218 S-----QLSPGQSOVQTTTSQPIQPIHTSL--QIP--SOGQPOQSP----- 2256
Db      2023 SAGKAMTDPKPAVSDENKINTFVETPVQKLDLLGNLPGSKRQPTPKAKAEALDLYGFK 2082
QY      2257 ---QVQSTQTLSSGQTLNVSVSSSRPQLIQOQPOQVIAPV---QLQOOQVQLSQI 2309
Db      2083 ELFOTPGHTEESMTDDKITEVSKSPQESPKTSRSSKQRLKIPLVKVDKMEEPKPLAVSKL 2142
QY      2310 --SQSVVAQIOAQOQSGVPQIKL--QLPIQIQOQSSAVQTHQIQNVVT--VQAASVQEQLOQR 2364
Db      2143 TRTSGETTQTHTEPTGDSKSIKAFKESPKQIILDPAASVTGSRRLQTRKEKARALELD 2202
QY      2365 VQQL-----RDQOQK-----KKOQOIEKR-----EH 2386
Db      2203 FKELFAPGHTTESMTIDKNTKIPCKSPPPPELDTATSTRKCPTRPRKEVEBELSAVER 2262
QY      2387 TLQASNOSEITQKQVWKHNAVIEHLQO--KKSMTTPAEREENORMIVCNQVMKYILDKID 2444
Db      2263 LTQTSQGSTHTHKEPASDEG--IKVLQORAKKKPNVPEEESRR----- 2305

```



```
QY 2445 KEEQAAKKRRREESVEQKRSKONATKLSALLFKHKEQLRA-----ETLKKR 2491
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2306 -----RPRAPKEAQLDLAGFTELSETSGHTAGKATKIPCESPPLEVVDIT 2358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2492 ALLDKDLQIEVQELKRDILKIKKEDLMOLAQ-----ATA 2526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2359 ASTKRHLRTRVQR-----VOVKEPSAVKFTQTSGETTADKEPAGEDGKIKALKESAKQ 2413
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2527 VAAPCPVTPVLPAPPAPPPSP-----PPGVQHTGLLSTPLTPVASQRRKREEEK 2578
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2414 TPAPAASTGSRRRPRAPRESAQAIEDLAGFKDPAAGHTPEESMTDDTKTKIPCKSSPELE 2473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2579 DSSSKKKKMISTTKETKDKTKLYCICKTPYDESKFYIGDRCONWYHGRVCVGLQSE 2638
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2474 DTATSSKRRPRTRAQVEVKEE---LLAVGK-----LTQTS 2506
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2639 AELIDEVVCQOSTEDATVLTPLTEKDYEGLRVLSLQAHKMAWPFLEVPDPN-DAP 2697
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2507 GE-----TTHYDKPEVGGKGTGA-----FKQPAKRNVDAE 2537
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2698 DYYGVIKERP-----DLATMEERVQR-RIYEKLTFFVAD 2730
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2538 DVIGSRRRQPRAPKEAQLDLASFQELSTQPGHTEELANGAAD 2581
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
T21435
hypothetical protein F26H11.3c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21435
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21435
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-452 <WIL>
A:Cross-references: EMBL:Z81515; PIDN:CAB54234.1; GSPDB:GN00020; CBSP:F26H11.3c
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CBSP:F26H11.3c
A:Map position: 2
A:Introns: 58/1; 89/3; 230/2; 290/2; 334/1
C:Superfamily: bromodomain homology
F;313-368/Domain: bromodomain homology <BRO>

Query Match 2.8%; Score 404.5; DB 2; Length 452;
Best Local Similarity 24.6%; Pred. No. 2.6e-08;
Matches 105; Conservative 82; Mismatches 159; Indels 81; Gaps 10;

QY 2360 EQLQR-VQOLRQOQKKQOQIEI-----KREHTLQASNQSEIIQKVVMKHNAVI--- 2409
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 ERLKKEIKRRTWENEAQOCCGLTPWKRARPHRAAKPAEYKKEVINPADITLGGD 80
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2410 --EHLKOKKSMTPAEREENQIVCNQVMKYIILDKIDKEEKQAKKRKEESVEQKRSKQ 2467
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 81 TYDYVVEQKPTESIAATNVSRRTTSANLSKSEDDR-DKPEQSTQAPKSKERRTSEPPASH 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2468 NATKLSALLFKHKEQRAEILKKRALLDKDLQIEVQEEELKRDILKIKKEDLMOLAQATV 2527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 VAFHTPGSATPHDINLSIEHCTCQKTFDAS-KLYIQCEL---CARYWHGDCVGAEQTIL 195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2528 AAPCPVTPVLPAPPAPPPPPPGVQHTGLLSTPLTPVASQRRKREEEKOSSSKSKK 2587
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 -----GLEHW-----SCECTEEQER----- 211
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2588 KMISTTSKETKDKTKLYCICKTPYDESKFYIGDRCONWYHGRVCVGLQSEAEILDEYVC 2647
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 -----VKQOPALYCVQKPYDPTKFYVVGDCSQGFHPCVGTTRAEAAQADYNC 262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2648 PQC-----QSTEDAMTVLPTLTKDYEGLRVLSLQAHKMAWPFLEVPDPNDA 2696
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
}
```

```
Db 263 PACTREAEGYESDASVSGSSRVSVQLTRADYTHVFELLELELLEHRMSTFPRPVDUNEF 322
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2697 PDYGVGKPEMDLATREVRQRYEKLTFEVDAMTKIFONCRVYNSDSDFPQCAEVL 2756
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 PDYEKFKKMDLSTTKKVERTEYLYLSQFVDVNDQMFENAKTYNPKGNVFKCAETMQ 382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2757 SFFVQKL 2763
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 EVDKKL 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: EMBL:X65550; NID:g415818; PIDN:CAA46519.1; PID:g415819
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F;29-91/Domain: Kinase interaction domain homology <KIH>

Query Match 2.8%; Score 404.5; DB 2; Length 3256;
Best Local Similarity 18.6%; Pred. No. 3.1e-07;
Matches 546; Conservative 378; Mismatches 1049; Indels 955; Gaps 136;

QY 339 IIEEDTENENKKIWIYSTKVQLAELIDCLDK-----DYWEAELCKLLEEMREIHRHMDI 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 VLPTEIHNPEFLTLWLATQVERKIQKDSLSKPEKLTGTAGQCMCSGLPLGSS-----VDI 486
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 395 TE-DLTNARG-----SNKSELAANEEILE-----SIRAKGDDIDN-----VKSPEE 436
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 NNFGDSINESEGIPLKRRRVSGFGLRPELFDENLPNTPLKRGAPTCKKSLVMHTPPV 546
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 437 TEKDKNETENDSKDAENRE---EFEDQSL-----EKDSDDKTP-DDDPEQKSEVGDfKS 488
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 547 LKKIIEQPQPSCKQESGEIHVEVKAQSLVSPAPSPKPTPVASDORRSCKTAPASS 606
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 489 EKNGELSESPGAKGASGSTRIITRLRNPD-----KLSQLKSOQVAAAHAANKLFK 542
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 SKSQTEVPK-----RGGERVATCLQKRVSISRSQHDILQITCSKRRSGASEAN--- 654
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 543 EGKEVLVNSQGEISRLSTKK---EVIIMG---NINNYFKLGOEGKRVY--YHNOYSTN- 593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 655 ----LIVAKSWADVVLGAKQOTQTVIKHGQPSRMNKRQRRPATPKPVGEVHSQSFSTGH 710
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 594 -----SFALNKHQHRDHDKRR--HLAHKFCFLTPAGEFK-----WNGSVHGSKVL 637
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 ANSPCTIILCAKTEKHVHPARYRVNLNFIISNQKMDFKEDLSIAEMFKTPVKEQPOLT 770
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 638 ISPLRTITQLENNISPF-----LHPNASHRANWIKAVQMCCK--PRFALALAI 687
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 771 -STCHIAISNELLKGQFGQTDGEBPLPTSESFGNVPFFSAQNAAKQPSD----- 822
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 688 LECAPKPVMLPIWREFLIGHT-----RLHRWTSIERE----- 719
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 823 -KCSASP----PLRRQCIRENGNVAKTPRNTYKMTSLETSTSTETETPSKTVTVNSGR 877
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 720 --EKEVKVK--KEKKOEEETM-----QOATWVKYTFPVKHQWVKQKGEYRTGY 766
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 878 STEFRNIQKLPVESKEEETNTEIVECILKRGQKATLLQO-----REGENKEI--- 925
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 767 GGSWISKTHVYRFVFKPLPGNTNVNRYKSLSEGTNNNDENMDESCK-----RKCS--RSP 819
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
}
```

```

Db      926  -----ERPET-----YKENIELKEN--DEKMKAMKRSRTGWQCAPMSDL 964
QY      820  KKIKIEDSE--KDEVG-----SDAAKADQNMEDISKITEKKDDVKEILDSDSKP 871
Db      965  TDLKSLPDTELMDKTARGONLQOTDHAKAPKSEKGIKTKM-----P 1006
QY      872  CKE--EPMEVDDDMKTESHVNCQESSOVVVNVSE-----GFHLRTS----- 911
Db      1007  COSLQP-----EPINPTHTKQQLKASLGKGVVKEELLAVGKFTSGETHTHREPADG 1062
QY      912  -----YKKTKY-----SSKLDGLERRIKOFTLEEKORLEKIKLEGGIKIGKISTNSSK 961
Db      1063  KSIRTEKESPROILDPAARVTGM--KKWPRTPKEEAQSLDL--AGFKELFOTPGPSEE 1117
QY      962  NLSESPVITRAKECQSDSRMQEOSPNANDOPE-DLIQCSQSDSVLRKMSOPSHHTNK 1020
Db      1118  SMTDEKT---TKIACKSPPPSVDTPTSTKOWPKRSRUKADVEEFALRLKLTPSAGKAM 1174
QY      1021  LYPKDRVLDDYSI-----RSPETKCPKONSTEN-----DIE 1051
Db      1175  LTPPAGDEKDIKAFMGTVPQKLDLAGTLPKSKRLOLTPKEKAQALDLAGFKELFQTP 1234
QY      1052  EKVSDLASRGQ-----EPTKSKTKGNDDFFIDDDSKLASADDIGTLI--CKNK 1095
Db      1235  GHTEELVAAGTKTKIPCDSPQSDPDVDTPTSTKOR-----PKRSIRKADVEGELLACRNL 1288
QY      1096  KPLIOEESDITIVSSSKSALHSSVPKSTNDRD-----ATPLSRAMDFEGKLGCDSESNSTL 1150
Db      1289  MP-----SAGKAMHTPKPSVGEEDKIIIFVGTVPQKL-----DLT 1323
QY      1151  ENSSDTVSIODSSEEDIVQNSNESISEQFRTREQDVEVLEPLKCELVSGESTG--NCEDR 1209
Db      1324  ENLTGSKRRPQTPKEEAQALDLAGFKELFOTPGHTEEA-----VAACKTKMPCCESS 1376
QY      1210  LPVKGTEANGKKPSQOKLEBRPNVKCSDQIKLANTDDKKNNENRESEKKGQRTSTQIN 1269
Db      1377  PPESADTPTSTRROPKTPLEKRDYQKELSALK-----KLQTSGETHTHTDKVP 1424
QY      1270  GKDNKPKIYAGECLKEITESRVSNGV--EPKVNINKIIPENDIKSLTVKESAIRPFIN 1328
Db      1425  GGEDKSIKNAFRETAKQKLDPAASVTGSKRHPKTK--EKAQLEDLAGW--KELFQTP--- 1477
QY      1329  GDVIMEDERNSETSKHLLSSSDAEGNYRDSLETLPSTKESDSTOTTPSASCPSNS 1388
Db      1478  ---VCTDKPTTHXKTKIACRSQPD-----PVDTPSSKQSKR 1513
QY      1389  VNQVEDMEIETSEVKKVTSS-----PITSEESNLSNDFIDENCLPINK---NENV 1436
Db      1514  SLRKVDVEEFPALRKRTPSAGKAMHTPKPAVSKEKNIYA--FM---GTPVKLDLTENL 1568
QY      1437  NGESKRKTVITEVTMTSTVA-----TESKTVIKVEKGDQTVVVSSTENCAKS 1484
Db      1569  TGSKRRLQTPKEKAQALDLAGFKELFOTRGHTEESMT---NDKTAKVACKSSQPDLDKN 1625
QY      1485  TVTTTTTTVTKLSTPSTCGSVDDIISVKQDSKTVTTTVDLSLTGTLTMTVSKBYS 1544
Db      1626  PASSKRLKTSIG--KVGKBEELLAVGLQTSGETHTHTTEPTGDKGSKMAFAMESPKQI 1683
QY      1545  TRDKVKLMKFRPKTKTRSGTA-----LPSYRK--FVTKTSKSIIFLVPND 1588
Db      1684  LDSAASLTGSKRQLRTPKGVKEVPEDLAGFELFQTPSHTKESMTNEKTKTVSRASQPD 1743
QY      1589  L-----KKLARGGIREVPYFNYNKAPALDIMPYSPRPTFGITWRYRLQTVKS 1637
Db      1744  LVDTPSSKQPKRSLRKADTEE-----EFLAKRQKTPSAG---KAMHTPKP 1787
QY      1638  LAGVSLMLR--LWASLRWDDMAKVPPGGSTRTSETTEITTEILIKRRDVGYPGIR-- 1694
Db      1788  AVGEEKDINTFLGTVPQKLDOPGNLP--GSNRLQTRKEKAQALFELT-----GFREL 1838
QY      1695  -----FEYCIRKILIC--PIGVP--ETPKETPTPQKGLRSS-----ALRPKRP 1733

```

```

Db      1839  FOTPCNDNPADEKTTKKILCKSPQSDPADTPTNTKQRPKRSLLKADVEEFLAFKRLTP 1898
QY      1734  -----ETPKOTGPVLIETWAAEELELEWIRAF-----AERVE-----KE 1768
Db      1899  SAGKAMHTPKAA-----VGEK-----DINTFVGTPEVKLDLLGNLPGSKRRPQTPKE 1946
QY      1769  KAAVEQQAARLEOQKQPVVIATSTTSPTSSTSTISPAQKVMVAPISGSVTTGKMWLT 1828
Db      1947  KAKALEDLAGFKELFOTPGHTEESMTDDKITTEVSKSPQDPVKTP-----TSSKORLQ 2000
QY      1829  TKVGSFATVTFQONKNFHQTFATVWQOGSNQSGVQVQOKVL--GIIPSTGTSTQOTETS 1886
Db      2001  ISLG-----VGKKEEVLPGVKLQOTSGKTQOT--- 2028
QY      1887  FOPRTATVITRPNTSGSGGT---TSNSQVI---TGPQIRPGMTVIRTP-----L 1929
Db      2029  -----HRETAGDGKSIKAFKESAKQMLDPANVGTGMERWP-----RTPKEAQSL 2073
QY      1930  QOSTLGAIIIRTPVMVQPGAPQOVMTQIIRGQVPSTAVSAPNTVSTPGOKSLTSATSIS 1989
Db      2074  EDLAGFKELFQTPDHTEESTTDDKTKI-----ACKSPPPESMDTPTST- 2117
QY      1990  NIOSSASOPRPOOGOVKLT--NAOLTQLTQGHGNGOGLTVVIOGOGQTTGQLQLIP--- 2044
Db      2118  -----RRPKTPLGKRDIVEELSALKQLTO-----TTHTDKVPGE 2153
QY      2045  QGVTVLPGGOQLM-----QAAMPNGTVQ-----RELFTPLATTATTAS 2083
Db      2154  DGINVFRETAKQKLDPAASVTGSKRQRPKGAQPLEDLAGLQKELFQTPVCTDPTTH 2213
QY      2084  TTTTIVSTTA-----ACTGE--QROSKLS--POMQVHQDKTLPQAQSSVGPKAQOTA- 2134
Db      2214  EKTTKIACRSPDPVGTPTTIFPKQSKRSRUKADVEEESLALRRKTPSVGKAMDTPKPAG 2273
QY      2135  -----QPSARPOQTOPQSPAQPEVQTPQVQTTVSS-----HVP 2171
Db      2274  GDEKDMKAFMGTPVQKLDLPGNLPGSKRWQTPKE-----KAQALDLAGFKELFQTP 2326
QY      2172  SEAOPTHAQSKKPOVAQAQSPQSNVQOSQSPVVSQSPQTRIRPS----- 2215
Db      2327  GTDKPT--TDEKTKIACKS--PQ-----PDPVDTFASTKQRPKRLKADVEEFLALR 2377
QY      2216  --TPS-----QLSPGQOSQVOITTSQPIPIQHTSL--QIP--SOGOPOSQP--- 2256
Db      2378  KRTPSAGKAMDTKPAVSDENKINFTVETPVQKLDLLGNLPGSKRQTPQTPKEAEALD 2437
QY      2257  -----QVQSSQTSLSSQTLNQVSVSPSPQLQIQPOPOVIAVP-----OLQOQOVQ 2305
Db      2438  VGFKELFQTPGHTEESMTDDKITTEVSKSPQSPESFKTSRSKQRLKPLVKVDMKERPLA 2497
QY      2306  LSQI--OSQVVAQIAQOOSQVPOQIKL--OLPQIQOSSAVOHTQIONVVT--VQAASVQE 2360
Db      2498  VSKLTRSGTGTQHTTEPTGDSKSIKAFKESKQILDPAASVTGSRRLQTRKEKARALE 2557
QY      2361  QLORVQOL-----RDQOQK-----KQOQOIEIKR--- 2384
Db      2558  DLVDFKELFAPAGHTEESMTIDKNTKIPCKSPPELTDATSTKRCFKTRPREKVEELS 2617
QY      2385  --EHTLAQSNQSEIIQOVVKNHNAVIEHLQ--KKSMTPAEREENORMIVCQVMKIYL 2440
Db      2618  AVERLTQSGSTHTHEKAPASGDEG--IKVLKORAKKPNVEEPPSR--- 2664
QY      2441  DKIDKEQAQAKKRRKRESVEQKRSQONATKLSALLFKHKEOLRA-----EI 2487
Db      2665  -----RPRAPKEAQPLEDLAGFTELSTSGHTQESLTAGKATKIPCESPPLEV 2713
QY      2488  LKRALLDKDLQIEVQBELKRDLIKKEKDLMLQALQ----- 2523
Db      2714  VDTTASTKRLHRLTRVQK-----VQKEPSAVKFTQTSGETTDADKEPAGEDKIGKALKE 2768
QY      2524  -ATAVAPCPVTPVLPAPPAPPPSP---PPGVQHTGLLSTPTLPVASQKRKR 2574
Db      2769  SAKQTPAPASVTGSRRRPRAPRESAQAIEDLAGFKDPAAGHTEESMTDDKTTKIPCKSS 2828

```


Qy 1241 KLNKTTDKNNRESEKKGORTSTFQINGKDNKPKYILKGECLKESRSRVVSGNVEPK 1300
Db 297 KKTTSKTKCTKTTTPVPPSSSTTE-----SSAPVTPPSSSTTESSAPVTSSTTE-- 350
Qy 1301 VNNINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSSTKSHLLSSSDAENYRD 1360
Db 351 --SSAPVP--TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP 394
Qy 1361 SLELTPSKESDS---TQTTTPASCPESNVQVEDMEIETSEVKVTSSTPITSE---- 1413
Db 395 VPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV-TSSTTESSAPVTSSTTES 453
Qy 1414 -----EESN--LSNDFIDENGLPINKNENNGESKRKTVITEVTMTST-VA 1457
Db 454 SSAPVTPPSSSTTESSAPVTSSTTESSAPVTPPSSSTTESSAPVTSSTTESSAPVP 513
Qy 1458 TESKTVIKVBKGDKQTVVSTENCASKSTVTTTTVTKJSTPSTGSGVDIISVKEQSKTV 1517
Db 514 TPSSSTTESSAPAPTSSSTTESSAPVTSSTTESSAPVTPSS-----STTESSTP 568
Qy 1518 VTTVTDSLATTGGTLVTSMTVSKE--YSTPDKVKLMKFKRPPKTRSGTALPSYRKPVTK 1575
Db 569 VTSSTTESSAPVTPPSSSTTESSAPVTPPSSSTTESSAPAPTSSSTTESSAPVTS 628
Qy 1576 STKKSIFVLPNDLKLARKGIREVPYFYNNAKPALDIWPPSPRPFTGITWYRLQTV 1635
Db 629 STTESSA-----ETPKETPTPKGLRSSAL-----TTE 647
Qy 1636 KSLAGVSLMLRLWASLRWDDMAAKVPPGGGSTRTESEITIT-----TEIKR 1685
Db 648 SSSAPVTP-----SSSTTESSAPVTPPSSST-TESSAPVTSSTTESSAPVTSSTTE 701
Qy 1686 RDVGPYGRFEYCIRKICIPGV-----ETPKETPTPKGLRSSAL----- 1728
Db 702 SSSAPVTPPSSSTTESSAPVTPPSSSTTESSAPVTPPSSSTTESSAPVTSSTTES 761
Qy 1729 ----RPRRPETPKGTGVIIETVAAEELEWEIRAFERAEVEKEAQAQBQAKKLEQQ 1784
Db 762 APVTPPSSSTTESSAPVTPPSSSTTESSAPVTPPSSSTTESSAPVTPPSSSNTISS 821
Qy 1785 KP--TVTATSTTS-----PTSTTTSTISPAQKVWVAPISGVSVTTCKKVLTKVGS-- 1833
Db 822 APSSTPPTSSSTTESSVTPPSSSTTESS-----APVSSSTTESSAPVTPPSSSN 874
Qy 1834 ----PATVTFQONKNEHQFATVWKQGNNGSVVQVQKVLG-----IIPS----- 1875
Db 875 ITTSAPSSIPSSSTTESFST-GTIVTPSSSKYPCSQETSVSSTETITVPTKTTTSVTT 933
Qy 1876 -----STGTSQ--QFTTSFQPRATVTRPNTSGSGGTTNSQVITGPOIRPGM 1922
Db 934 PSTTTITTTVCGTGNAGETISGCSPKVTVTV-PTTTTTSVTSSTTIT-----T 985
Qy 1973 TVIRTPQ--QSTLQ--KALIRT-PVMVOPGAQQVMTQILRGQPVSSTAVS-----AP 1970
Db 986 TVCGTGNAGETISGCSPKVTITTVPCSTSPSTASESTTSTPTTVTVVSTTVTTE 1045
Qy 1971 NTVSSTPGQKSLTSATSTSNIQSS--ASQPPRPOQGV-KILMAQLTQLTQHGNGQGLT 2027
Db 1046 YSTSKPGGE-ITTFVTKNIPITVLTIAFTPSVTVTNFTPTTITTTVCSTGTSNA-- 1102
Qy 2028 VVIQCGQTTQQLQLIPGVTVL-----PGCQQLMQAAMPNGTVORFLPTLATATAS 2083
Db 1103 -----GETTSGCS--PKVTVTTVPCSTGTGXTTEA-----TTLVTTA--- 1138
Qy 2084 TTTTTSVSTAAGTGEORQSKLSPQMOVHODKTLPPAQSSSVGPAKAQPTAQSARPOQ 2143
Db 1139 -VTTTVTSTGTSNAGKTTTG---YTHKSVPTTVTTLAPS----- 1178
Qy 2144 TQPOSAPQEVQTOPEVOTQTVTSVSHVPSEAQPTHAAOSSKPVQAAQSQPOSNVQGSFVR 2203
Db 1179 ----APVTPPATNAVPTTITTTTECSAATNAAGETTIVSASAKTIVSSAGENTAPSAT--- 1231
Qy 2204 VQSPSQTIRPST---PSQLSPGQSQSQVQTT---TSQPIPIQPHLSLQIPSQGQPOSQPV 2258

Db 1232 --TPVTTAT-PTTVITTESSVGTNSAGETTTGYTTKSIPTTYITTL-IPGSNGAKNYETV 1287
Qy 2259 QSTSTQLSSGQTLNQVSVSSPRPQLOIQPOQPVQVLAQVQLSQVLSQIQSVVVAQIQ 2318
Db 1288 ATAT-----NPISIKTTS-----OLATTASASSVAPVV 1315
Qy 2319 AQOSGVPQIKIQLPIQIQSSAVQTHQIQNV-VTVQAAS 2357
Db 1316 TSPS-----LTGFLQASGSNAVATYVSPSISSTVQGA 1348
RESULT 15
T21433
hypothetical protein F26H11.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21433
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21433
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-405 <WIL>
A:Cross-references: EMBL:Z81515; PIDN:CAB04198.1; GSPDB:GN00020; CESP:F26H11.3a
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.3a
A:Map position: 2
A:Introns: 11/1; 42/3; 183/2; 243/2; 287/1
C:Superfamily: bromodomain homology
F:266-321/Domain: bromodomain homology <BRO>
Query Match 2.8%; Score 399.5; DB 2; Length 405;
Best Local Similarity 39.4%; Pred. No. 3.4e-08;
Matches 71; Conservative 35; Mismatches 63; Indels 11; Gaps 1;
Qy 2595 KETKDKTKLYCICKTPYDESKFYIGDCRQNMVYHGRVGLQSEAEILIDEYVCPQC----- 2650
Db 163 ERVKDQPALYCVCKPYDDTKFYVGCDSQGVFPCVGTTRAEABQAAADYNCPACTREA 222
Qy 2651 --7-----QSTEDAMTVLTPLTEKDYGLKRVLSLQAHKMWPFLEPVDNDAPDIYGV 2703
Db 223 EGYESEASDVSGSSRVSVQLTRADYTHVFELLELLEHRMSTPFRNPVDLNEFPDYKFI 282
Qy 2704 KEPMDLATMEERVORRYEKLTEFVADMTKIFDNCRYNPNSDSPFYQCAEVLSEFFVQKL 2763
Db 283 KKPMDLSTITTKVERTEYLYLSQFVNDVNMQMFENAKTYNPKGNVFKCAETMQEVEDKKL 342
Search completed: September 24, 2003, 01:18:40
Job time : 59.3354 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2003, 19:30:01 ; Search time 19.068 Seconds
(without alignments)
6858.664 Million cell updates/sec

Title: US-09-698-295-10
Perfect score: 14333
Sequence: 1 MVSEEEEDGDAETQDSE.....KLKGFKASRSHNNKLOSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	3363	23.5	1 FALZ_HUMAN	Q12830 homo sapien
2	441	3.1	1 TRX_DROME	P20559 drosophila
3	440.5	3.1	1 BAZA_HUMAN	Q9uif9 homo sapien
4	437	3.0	1 PCLO_CHICK	Q9pu36 gallus gall
5	421.5	2.9	1 PCLO_HUMAN	Q9y6v0 homo sapien
6	419.5	2.9	1 BAZA_MOUSE	Q9lye5 mus musculus
7	404.5	2.8	1 K167_HUMAN	P46013 homo sapien
8	401.5	2.8	1 AMYH_YEAST	P08640 saccharomyc
9	396.5	2.8	1 BAZB_CHICK	Q9dei3 gallus gall
10	391	2.7	1 MUC2_HUMAN	Q02817 homo sapien
11	390.5	2.7	1 PCLO_MOUSE	Q9qyx7 mus musculus
12	388	2.7	1 PCLO_RAT	Q9jks6 rattus norv
13	385	2.7	1 BAZB_HUMAN	Q9uif8 homo sapien
14	382	2.7	1 NCR1_MOUSE	Q60974 mus musculus
15	380.5	2.7	1 AKA9_HUMAN	Q99996 h a-kinase
16	379.5	2.6	1 PHP_DROME	P39766 drosophila
17	375.5	2.6	1 ANK2_HUMAN	Q01484 homo sapien
18	374.5	2.6	1 MYS2_DICDI	P08799 dictyosteli
19	374	2.6	1 ANK3_HUMAN	Q12955 homo sapien
20	367.5	2.6	1 GIAN_HUMAN	Q14789 homo sapien
21	364	2.5	1 UN99_CAEEL	Q01761 caenorhabdi
22	360	2.5	1 MLP1_YEAST	Q02455 saccharomyc
23	360	2.5	1 ATRX_HUMAN	P46100 homo sapien
24	358.5	2.5	1 TRX_DROVI	Q24742 drosophila
25	356.5	2.5	1 NCR1_HUMAN	Q75376 homo sapien
26	355	2.5	1 HFC1_HUMAN	P51610 homo sapien
27	354.5	2.5	1 TPR_HUMAN	P12270 homo sapien
28	353.5	2.5	1 MUSB_HUMAN	Q9hc84 homo sapien
29	348.5	2.4	1 Y956_YEAST	Q04893 saccharomyc
30	347.5	2.4	1 MAPB_RAT	P15205 rattus norv
31	343	2.4	1 MAPB_MOUSE	P14873 mus musculus
32	342.5	2.4	1 BPAL_HUMAN	Q94833 homo sapien
33	337.5	2.4	1 USO1_YEAST	P25386 saccharomyc

RESULT 1	FALZ_HUMAN	STANDARD;	PRT;	810	AA.
ID	FALZ_HUMAN	Q12830;			
AC	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Fetal alzheimer antigen (Fetal Alz-50-reactive clone 1).				
GN	FALZ OR FAC1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fetal brain;				
RX	MEDLINE=95347245; PubMed=7621746;				
RA	Bowser R., Giambrone A., Davies P.;				
RT	"FAC1, a novel gene identified with the monoclonal antibody Alz50, is developmentally regulated in human brain.";				
RL	Dev. Neurosci. 17:20-37(1995).				
CC	-! SUBCELLULAR LOCATION: DETECTED IN BOTH THE CYTOPLASM AND NUCLEUS OF CELLS IN THE DEVELOPING CORTEX. IN THE ADULT BRAIN, IT WAS SEEN ALMOST EXCLUSIVELY IN THE NUCLEI OF NEURONS OF THE NEOCORTEX. IN THE BRAINS OF ALZHEIMER DISEASE PATIENTS, THE PROTEIN IS LOCALIZED IN A SUBSET OF AMYLOID-CONTAINING PLAQUES.				
CC	-! TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THE FETAL BRAIN. EXPRESSION IS MUCH LOWER IN ADULT BRAIN AND IS HIGHER IN NEURODEGENERATIVE DISEASES.				
CC	-! SIMILARITY: Contains 1 PHD-type zinc finger.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; U05237; AAA97522.1; -				
CC	PIR; G01252; G01252.				
DR	TRANSFAC; T04682; -				
DR	Genew; HGNC:3581; FALZ.				
DR	MIM; 601819; -				
DR	GO; GO:0005737; C:cytoplasm; TAS.				
DR	GO; GO:0005634; C:nucleus; TAS.				
DR	GO; GO:0007399; P:neurogenesis; TAS.				
DR	InterPro; IPR004022; DDT_dom.				
DR	InterPro; IPR001965; Znf_PHD.				
DR	Pfam; PF02791; DDT; 1.				
DR	Pfam; PF00628; PHD; 1.				
DR	SMART; SM00571; DDT; 1.				
DR	SMART; SM00249; PHD; 1.				
DR	PROSITE; PS01359; ZF_PHD_1; 1.				
DR	PROSITE; PS50016; ZF_PHD_2; 1.				
KW	Nuclear protein; zinc-finger.				

RESULT 2
TRX_DROME
ID TRX_
AC P206
DT 01-0

Db 1771 SAEVFEID-----GSLQNHSAVARGMIKCTVCGNRGAT-----VCCNVRSCG 1815
 Qy 1147 ---NSTLENSDTSVSIQDSS-----EDMIVQNSNESISEQFRTREODVEVLPLKC 1195
 Db 1816 EHYHPCARSIDCAFLTDKSMYCAHAKGNALKANGSPSV-----TYESNFEVSRPVV 1870
 Qy 1196 ELVSGESTGNCEDRLPVKGTENGKPSQOK-----KLEER-----PVN 1234
 Db 1871 EL-----DRKKRKLIE-----PARVQFHIGSLVLRQLGAIVPRFSDSYEA VVPIN 1915
 Qy 1235 -KCS-----DOIKLKNITDKNNENRESE-----KKGQRTSFTQINGRDNKPKIYLKGE 1282
 Db 1916 FLCRLWLSKPEWKIVEYIVRTTIONSSTLALDVGRTYVDHTNPNNSKEYVOLGMAQI 1975
 Qy 1283 CLKEISESR---VVSQNVGE-----PKVNNINKIIPENDIKSLVAKESAIRPFINGDVIME 1334
 Db 1976 ARWHTSLARSEFLNGGTGDSGEFPNPSNC---VPPDQNTTEEPQOQADLLPPELKDAIFE 2033
 Qy 1335 DFNERNSETKSHLLSSDAEGN---YROSLETLPSTKES-DSTQTTTPSASCPSNSVN 1390
 Db 2034 DL-----PHELLDGISMLDIFLYDDKDLFAISEQSKDGTQAMTS-----N 2074
 Qy 1391 QVEDMETETSEVKKVTSSPITSEESNLSDNDFIDENGLPI-NKNEINVNGESKRBKTVITEV 1449
 Db 2075 QAQNQOQAGANSVS-----ICDEDTRNSNTSLG-NGWPASNPVEDAMLGAARN---SSQ 2126
 Qy 1450 TTTTSTVA-----TESKTVIKVEKGDQKQTVVSSSTENCACAKSTVTTTTTITVKLSTPSTG 1502
 Db 2127 VOMLKTLAWPKLDGNSAMATAIKRRKLSK-----NLAEGVFLTLSSQORNKEMATV 2178
 Qy 1503 GSDIISVKESQKTVVTTVDSLTITGGHIVTSMT---VSKEYSTRD-----KVKLMK 1553
 Db 2179 AGV-----SRQOSISETSEGVATTSGSVRSKSFVWSAAKRYFEKSEGREEAAMKRMQ 2232
 Qy 1554 FSRPKKTRSGTALPSYRKFTKSTKSIFFVLNDLKLARKGGIR-----EVFYFNNAK 1609
 Db 2233 MDG-----VDDSITEFRIIISGDGNLSTAFSGQVKCDRCQCYRNWDAF 2276
 Qy 1610 PALDIWYPSRPRTFGITWRYLQTVKSLAGVSLMLLLWASLRWDDMAAKVPPGGGSTR 1669
 Db 2277 QR-----HLPSCSPT-----MSSNETSDVS-----GQGMIN 2303
 Qy 1670 TETSETITTEIKRQVGPYGRFEYCKIRKICPIGVETPKETTPQR----- 1720
 Db 2304 NATQISAESLNELOQLLANAGLNY-----LOSATSFPOVQRLGSLGQFGL 2350
 Qy 1721 KGLRSSALRPK-----RPETPKQTPVLIETWVAEELELW---ETRAFAERV-- 1765
 Db 2351 QGLQLOLQPSLNGEFLSOPNPTQANT-----DDLQIYANSLOGLAANLGG 2399
 Qy 1766 -----EKEKAQAVEQOAKRLEQOKPVTIATSTTSPTSTSTTSISPAOKVM 1811
 Db 2400 GFTLAQPTVTPAPAQOLIAVSTNPDGTQQFTQIPQTMQATTTPTATYQTLQATNTDKIM 2459
 Qy 1812 VAPISGSVTTGKVLIT---IKVGSFATVTPQONKFNHOTATWVK-----QGQNSGVVQ 1864
 Db 2460 L-----PLTAAGKPLKTVAKAAQAAVAKORQLKSHGVKPIQAKLPHQPQHQOQOQTQ 2514
 Qy 1865 VQQ---KVLG-----IIPSSTGT-----SQQTTSF----- 1887
 Db 2515 VQQPTVWGQNLQOLLFQSSTQTPAQIILPAQOQNIISFVYTGDSQOQPLOYISIP 2574
 Qy 1888 -----QPR-TATVTRPNTSGSG-----TTSNS---QVITGQPIRGM 1922
 Db 2575 TAGYKQPQPTPTFTLTAPGAGATYLTQDASGNLVLTTTPSNSGLQMLTAQSLQAOP 2634
 Qy 1923 TVIRTPLOOSTL-----GKAIITPVMVQPGAPQVM----- 1956
 Db 2635 QVIGTLIQPQIQLGGGADGN-----QPSNQOPLILGGTGGSSGLEFATTPSQV 2695
 Qy 1957 IIRGQPV-----STAVSAPNTVSST-PG-----QKSLTSATSTSIQISASQPPRPQ 2003
 Db 2686 ILATQPMYGLLETTIVQNTVMSQQQFVSTAMPGLMSQNASFSATTTQVFQASKIEP----- 2740

Qy 2004 GQVKLTMAOLTQITQHGNGOGLTV-----VIOQGGQ---TTGQLOLIPQGVTVLPGGQOL 2057
 Db 2741 -IVDLPGAVVVLNNTGDASSAGTFLNAASVLAQOQODDTTQI----- 2782
 Qy 2058 MQAAMPNGTVQRELFPTPLATTATTASTTTTSTTAAGTGEQROSKLSPOMOVHODKTL 2117
 Db 2783 ----LQANAFQ-FQSVPSTSGASTSMDYTSPVNVTA-----KIPVTOIKRTNAQA 2828
 Qy 2118 PAQS-SSVGPAPAKAQO-----TAQPSARPQPTQOPSPAQPEVQTOPEVOTOTVSSH 2169
 Db 2829 KAAGISGVKVPPOPVVNVKVLPTSIVTQOSQOVQVKNLKSQVYKGAASGTGTTGCGA- 2887
 Qy 2170 VPSEA-----OPTHAASSKEPV-----AASQPSNVQOGSPVR 2203
 Db 2888 PPSIASKPLQKKTMNIRPIHKLEVPKVMKPTPKVQONQNHSLQOQOQOQPOLQOQIPAV 2947
 Qy 2204 V---QSP-----SOTRIIRPST-----PSQLSPGOQSOVQVTTTSOPI---PIQPH 2241
 Db 2948 VNVQVPKVTISQORIPAQTOOQLOQAQMIHIPOOQOPLQOQOVQVPSMPIITLAEAPV 3007
 Qy 2242 TSLQIPSQGPQSQPOVQVSTQTLSSGQTLNVQSVSPSR----- 2281
 Db 3008 VQSQFVMEPQALEQELANRVQHFSSTSSSSSSNCSLPTNVVNPQQQAPSTTSSTTRP 3067
 Qy 2282 ----POLQIQPOP-----QVIAVPOLOQOVQ---VLSQIOSOVVAQIAQOQSGVPOQIK 2329
 Db 3068 TNRVLPQMOQREAPLNECPVVSSTPPKPEBPIIHOMTSASVSKCAQKSTLPSPV- 3126
 Qy 2330 LQLPIQIOSSAVOCTHOIQNVTVQAASVOELQVRQOQLRDQOQKQOQIEIKREHTLIQ 2389
 Db 3127 --YEALKSVSSVLES---IVPDMTDA--ILEEPVTVQSIYTEGLYKENSPEGSKTEQLLL 3180
 Qy 2390 ASQSEIIQVQVYKHNNAVIEHLKQKSM-TPAEREN-----ORMIVCN- 2433
 Db 3181 QQOQREQLNQLVNGVLLDKHTFQVEPMDTVYREEDLEEEDEDEDDFSLKMRTSACND 3240
 Qy 2434 ----QVMKYLIDKIDKE-----KQA 2450
 Db 3241 HEMSDSEEPVAVKDKISKILDNLNDDCADSIATATMEVDNADASGYQOMVEDVLTAAQS 3300
 Qy 2451 AKRKRKEESVEQKRSKONATKLSALLFKHKEQLRAETLKKRALID-KDLOIEVQEEELKRD 2509
 Db 3301 APTEEEFEGALETAEAATAATVINEADAH-----VLDLKLQQLNGVLELELR 3347
 Qy 2510 LKTIKKEKDLMLQAATAFAAPCPVPTVLPAPPAPPPPPPPPG 2553
 Db 3348 ---KEQRTVSQEQSKAAIIVP---TAAAPPEPPIQEPKMTG 3386

RESULT 3

BA2A_HUMAN
 ID BAZA_HUMAN STANDARD; PRT; 1878 AA.
 AC Q9UIF9; O005336; O15030; Q96H26;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bromodomain adjacent to zinc finger domain 2A (Transcription
 DE termination factor-I interacting protein 5) (TFI-I interacting protein
 DE 5) (Tip5) (hWALP3).
 DE BAZ2A OR TIP5 OR KIAA0314.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=20130112; PubMed=10662543;
 RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
 RT "A novel family of bromodomain genes";
 RL Genomics 63:40-45(2000).
 RN [2]


```
Db 103 SANPGSNLKDPPLLSQFSGG-----QYPLNGILGSRQPS-----PS 140
Qy 1016 HTTN-----KLY-----PKDRVLDVVSIRSPETCKPKQNSTIENDIEKVSJDLASRGQE 1063
Db 141 HNTNLRAGSQKFWANGTHSPMGLNFDQSLEYD---SFPDQN-----FEEVCS-----GTH 187
Qy 1064 PTKSKTKGNDFDDSKLASADDIGT-LICKNKKPLIOESDRIVS-----SSKSALH 1115
Db 188 PDEAAEKEMTSVVAEN-----GTGLVCSLE---LEEQPELKWCGNGSVSPVESLH 236
Qy 1116 SSVPKSTNDRDAPLSRAMDFEGLGCDSSNSTLENSDSTVSIQDSSSEEDMIVQNSNES 1175
Db 237 QEVSVLVPD---PTVSCLD-----DPSHLPDOLEDTPLS-----268
Qy 1176 ISEQFRTRQODVEVLEPKELCVSGESTG-----NCEDRLPVKTEANGKKPQSOQKL 1228
Db 269 -----EDSLEPFNSLAPVPVSGGLYGIDDTLMAEDKLP-----L 304
Qy 1229 EERPVKCSQOIKLKNITDKNNENRESEKKGORTSTFQINGKDNKPKIYLKGECLKEIS 1288
Db 305 EDSPVISALDCPSLNAT-----AFSLADDSQ-----332
Qy 1289 ESRVSVGNVEPKVNNINKIIPENDIKSLATVKESAIRPFINGDVIMEDFNERNSETSKSHL 1348
Db 333 -----TSTSFASPTSPVLGESVLQD-----NSFD-----L 359
Qy 1349 LSSSDAEGNVRDSLETLPSTKESDSTQTTTPSASCPSNSVNOVEDMEIETSEVKKVTSS 1408
Db 360 NNGSDAE---QEEME-----TQSSDFPSPSLTOPA---PDQSTIOLHP---ATSPAVSPTTS 407
Qy 1409 PITSEESNLSNDRIDENGLPINKNVNNGESKRKTIVITEVTMTSTVATESKTVIKVEK 1468
Db 408 PAVSLVSPASPEIS-----PEVCPAASTVV---SPAVESVVS 443
Qy 1469 GDKQTVVSSSTENCAKSVTTTTVTTKLSTPSTGSGVDIISVKREQSTVTTTTVTDLSLT 1528
Db 444 PASSAVLPAVSLVPLTASVTSKRASPVTPSPAAFPATSPANKDVSSFLETTADVEEITG 503
Qy 1529 TGGILVTSMTV-SKEYSTRDKVL---MKFSRPRKTSRGT-----AL 1566
Db 504 EGLTASGSDVWRRRIAPPEEVRPLPLQHWREVRIRKKGSHRWGETWYGPCGKRMKQF 563
Qy 1567 PSYRKFTVKSTKSFIFLPLNDLKLARKGGIREVPYFNYNAPALDIPWYPSRPTFGI 1626
Db 564 PEVIKLSRLNVHSV-----RRE-HFSF-----SPRMPVGD 593
Qy 1627 TWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAKVP---PGGSGTSTETSET-----1675
Db 594 FFEER-DTPEGLQ-----WVOLSABEIPSRIOAITGKRGRPRNTERAKTKEVPKVK 643
Qy 1676 -----BITTEILKRDVGYPYIGIRFYCIRKIIICPIGVPTPKETPTPQKGLRSSAL 1728
Db 644 RGRGRPPVKITELLNKTDNR-----LKKLEAQETLNEEDKAKIAKSKKKMRQKVQ 695
Qy 1729 R-----PKRETPKQGPVITETWVAEELELEWEIRAFARVEKEKAQAVEQ 1775
Db 696 RGECLTTTQGOARNRKQETSLAKHKAKKSKAEKEGKTQKQEKLEKVKRREKVKM 755
Qy 1776 QAKKRL-----EOKQPTVATSTTSPTSTST-----1803
Db 756 KEKEEVTAKPACKADKTLATQRLREKQKQOMILEEMKKPTEDMCLTDHQPDPFSRVP 815
Qy 1804 -----ISPAQKVMVAPI---SGSVTTGTM-----1825
Db 816 GLTLPAGAFSDCLTIVEFLHSFGKVLGPDFPAKDVPSLGLVQBLEGLICQGDLSLGEVQDLIVR 875
Qy 1826 VLTWKVGPATVTFQONKNFHOTFATWVKQGSN-SGVVQVQKVGLIIPSTGTSQ-OT 1883
Db 876 LLKAAHDPGFPSPYQCSQSLILGEKVSEIPLTRDNVSEILRCFLMAYGEVPALCDRLRTPQ 935
Qy 1884 FTSFQP-RTATVTIRPNTSGSGGT-----TSNSQVITGPQIRPGMTVIR 1926
Db 936 FOAQPPQKAAVLAPFVHVLNGSTLIINEIDKTLSEMSYRKNKWIIVESGLRR-----LK 990
1927 TPLOQSTILGKAIIRTPVMVQPGAPQVM-----TOIIRQOPYSTAVSAPNTVSSTPPGQKS 1981
Db 991 TVLAKRT-GRS-----EVENGRPEBCLGRRSSRIMEETSGMEEEESIAAIVPGR- 1042
Qy 1982 LFSATSTNIOSSASOPPRPQOQVKLTMAQLTQUTQHGNGQGLTVVLOGGQTTGQLQ 2041
Db 1043 --GRRDGEVDATASSIPELERQIEKLSKRQLFFRKLHLHSSQMLRAVSLGQDRYRRYW 1099
Qy 2042 LTP--QGVTVLPGCQOLMQAAMPNGT-----VQRELFTPLATTTASTTTTTSVITA 2093
Db 1100 VLPYLAGIPEVEGTEGNLPEEVIKKEITDSLKVAHAHSLNPALFSLMKMELAGSNTASSPA 1159
Qy 2094 AGTGEOROSKLSQPMQVHODKTLPPAQSSSVGPAKAQPTQAQPSARPQPTQTPQSPAQPE 2153
Db 1160 RARGRPRTK-----PGSMOPRHLKSPVRGDSQEQQAQLOPE 1197
Qy 2154 VQ-----TQPEVOTQTTVSSH---VPSEAQPTHAQSSKPOVA-----AOSQPOSNVQ 2198
Db 1198 AOLHAPAQPOPQLOLQLOKHGFLBQEGSPLSLGOSQHDLSQSAFLSWLSQTSOSLHLS 1257
Qy 2199 QSPVRVQSPQTRIRPSTPSQLSQSGOQOVQTTTQIP--TQPHTSLQIPSGQSQSQSP 2256
Db 1258 SS-VLTPDSSPKLDPA-PSQ--PPEPEPDEAESPDQALWENISQAQMPCNAAATPPP 1313
Qy 2257 QVQS-----STQTLSSGQTLNQVSVSSPSRPOQI-----QOPQ-PQ 2292
Db 1314 AVSEDOPTSPQOLASSKPMNRPSAANPCSP-VQFSSTPLAGLAPKRRRAGDFGEMPSPT 1372
Qy 2293 VIATPOLQOQ--VOVLSQISQSVVAQIQAOQSGVPOQ-----IKLQLP 2333
Db 1373 GLQOPKRRGRPPSKFKQMEQRYLTQLTAP--VPPMCGSWMWIRDPPEMLDAMLKALP 1430
Qy 2334 IQIQOSSA---VOTHQ--IQNVVTVQAAS---VQEQLQRVQ--LRDQOQKKQOQIEIKR 2384
Db 1431 RGIREKALHKLHNRDFLEQVCLRPSADPIPEPRQLPAFQEGIMSWSPKETYETDL-- 1488
Qy 2385 EHTLOASNOSEITOKQVWKNHNAVIEHLKOKKSMTPAEREENORMIVCQVMKYILDKTD 2444
Db 1489 -----AVLQWVELEQRYVIMSDLQIRGWTCPSPDSTREDLAYCEHLS-----D 1531
Qy 2445 KEKQAAKKRKEESVEQKRSKONATKLSALLFKHEQLRAELKKRALLDKDLQIE-VQ 2503
Db 1532 SQEDITWRGREGRLAPQRT-TNPLDLAVM-----RLAALEQNVERRYL 1576
Qy 2504 EELKRLDIKKEKDLMLQALQATAVAAPCPVTPVLPAPPAPPPPPPPGVQHTGLLSTP 2563
Db 1577 EPLWPTHEVVLEKALLS-----TPNGAPEGTTTETISEYETP 1612
Qy 2564 TLPVASQKRKEBKSSSSSKKKKMTSTTSKE-----TKKDTKLYCICKTPYDESKFYI 2618
Db 1613 RIRVWROTLER-----CRSAAQVCICLQGLERSTAWESVKNVTCVCRKGDNDFFLL 1665
Qy 2619 GDCRCQWNYHGRVCVILQSEAEILDE--YVCPQC--QSTEDAMT-----2658
Db 1666 LCDCCDRGCHLYC--HRPKWEAVPEGDFWCTVCLAQOQVEGEFTQKPGFPKRGKRSKY 1722
Qy 2659 -----VL-----TPLTEKDY--EGL--KR-----VLRL 2678
Db 1723 SLNFSEGDGRRRRVLLRGRESPAAGPRYSEEGSPSKRRRLSMNRHHSDJTFCEIILMEM 1782
Qy 2679 QAHKWAPLEPDPNDADPYGVVKEPMDLATMEERVQRRYFEKLETFEVADMTKIFDNC 2738
Db 1783 ESHDAAPFLEPVPNRLVSGYRRIKNPMDFTSMRERLLRGYTSSEFAADALLVFDNC 1842
Qy 2739 RYVNPSPSPFYQCAEVLESFFVQKLKGFKASRSIN 2773
Db 1843 QTFNEDDSEVKGAGHIMRRFFESWEFEOGKOAN 1877
RESULT 4
PCLO_CHICK
ID PCLO_CHICK STANDARD: PRT: 5120 AA.
```


AC Q9PU36; 3.0%; Score 437; DB 1; Length 5120;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Piccolo protein (Aczonin) (Fragment).
 GN PCLO OR ACZ.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99439764; PubMed=10508862;
 RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
 RA Killmann M.W.;
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
 RT zones, shares homology regions with rim and bassoon and binds
 RT profilin";
 RL J. Cell Biol. 147:151-162(1999).
 CC -!- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y19187; CAB60725.1; -;
 DR HSSP; P04110; 1A25.
 DR GO; GO:0045202; C:synaptic junction; ISS.
 DR GO; GO:0003509; F:calcium ion binding activity; ISS.
 DR GO; GO:0003544; F:calcium-dependent phospholipid binding acti. . . ; ISS.
 DR GO; GO:0003522; F:profilin binding activity; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001565; Synaptotagmin.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR PRINTS; PR00399; SYNAPTOTAGMN.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2_DOMAIN_1; 1.
 DR PROSITE; PS00004; C2_DOMAIN_2; 2.
 DR PROSITE; PS0106; PDZ; 1.
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
 KW Repeat.
 FT NON_TER
 FT DOMAIN 258 357 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT P-A-K-P-Q-P-Q-P-X.
 FT 2N_FING 368 392 C4-TYPE (POTENTIAL).
 FT 2N_FING 836 859 C4-TYPE (POTENTIAL).
 FT DOMAIN 2324 2343 POLY-PRO.
 FT DOMAIN 4414 4493 PDZ.
 FT DOMAIN 4627 4726 C2 DOMAIN 1.
 FT DOMAIN 5003 5094 C2 DOMAIN 2.
 KW SEQUENCE 5120 AA; 560751 MW; A658D9891B65B412 CRC64;

Query Match
 Best Local Similarity 17.6%; Pred. No. 4.9e-08;
 Matches 455; Conservative 385; Mismatches 945; Indels 798; Gaps 99;
 QY 415 EELESIRAKKGDIDNVKSPETEKKDKNETENDSKDAENRREFEQDSLEKSDDDTPDD 474
 DB 965 EKTLSADKIQGG-----IQKEDAKSKOGKLFKTPSADKIORVSKOEDSRLOOTKLTTPSS 1020
 QY 475 DPEQKSEYGVDFKSEKNGELSPGAGKSGASTRIITRLRNPDSKLSQKSSQVAAAA 534
 DB 1021 DXILHGQKEDIKFEA--KLAKIPSADK-----ILHRLQKEDFKLOQMK---MAKA 1067
 QY 535 HEANKLFKEGKEVLVNVNSOGEISRLSTKKEV---IMKGNINNVFKLGQEGKYRVYNQVS 591
 DB 1068 LSADKIQPEAQKEDVOLQEVRLSKAVSADKIQHGIQKDLNLQHVKTERTSSVEKIQEAO 1127
 QY 592 TNSFALKN-----HQ---HREDHDKRRHLAHKFCLTTPAGEFKWNGSVH 631
 DB 1128 ESKLQODKLPKLTSEDKIPATVSSDHKLLSKSEEDKPELLEK--STPHPKDK----- 1179
 QY 632 GSKVLTITSLRLTITOLENNIPSSFLLHPNASHRANWIK-----AVQMCSPREFALA 684
 DB 1180 KEQITAEITGHITQEKVEAECDKLH---EKKQEDVKKEDLTTCIPQWVKP----- 1229
 QY 685 LAILECAVPVVMPLIWRFLGHTLRLHMTSIEEREKEKVKKKEKEEETMQQATWTK 744
 DB 1230 ---EKAEEKTPVPV-----SRLPRSDHVE-AVREKIEKEDSDTSSTSSQOQ----- 1272
 QY 745 YTFPVKHQVWKQGEERYVTGGVSWISK-----THV---YRFVPLKPGTNNVNRKS- 795
 DB 1273 -----KSPQGLSDTGYSSDGISSLGSEIPSHIPSDEKDLPREPSQKDTISOESP 1321
 QY 796 -----LEGTKNNNDNN-----DE-----SDRKCSRSPPKIKIEPDSEK 830
 DB 1322 PPSDLAKLESTVLSILEAQASTLTDEKSVKRKELYETVSEQTKDQHKTKPLPVPESYS 1381
 QY 831 DEVKGSAAKADQNDMDISKITEKKDQVKKELSDSDKPKCKEPEME-VDDDMKMTESHV 889
 DB 1382 SDEEDLEATQEGERTTAADSKGASSTQDYKE-EDGNDTPARRQYSDVEDSESENSP 1440
 QY 890 NQESSQVDVNVNVEGFHLR-----TSYKKTKSKSLDGLLERRIK 930
 DB 1441 VPRKRRASVSGSSSDEYKRDSSQSGDEEDFTKIOIEMSADEDEAGSDEDFEIRNQLK 1500
 QY 931 QFTLEEKQRLKIKLEGGIKG-IGTKTSTNSKNLSPEVITKAKQCSQDSMRQ-----EQ 985
 DB 1501 EISVTSQKKEEVKSKA--KGTVGK-----HRRMARKSSAGYDEDAGRHSHWDD 1548
 QY 986 SPNANNQPEDLIQGCQSQSDSVLRMSDPSTHTNKLYPKDRLVDDVIRSPTKCPKQNS 1045
 DB 1549 DDETDESPEPKYRETKSQDGEELATSGG-----GLRRFKT-----IE 1587
 QY 1046 IENDIEEKVSLASRGQEPKSKTKGNDFIDDSKLASADDIGTLCKKNKKPLIQEESDT 1105
 DB 1588 LNSTITSKYSE-----TPQOKGILYFDEEPFL-----EMESLTDSPEDR 1627
 QY 1106 IVSSSKSALH-----SSVPKSTNDRDAPPLSRAMDFEGKLGCDSESNSTLNSSD 1155
 DB 1628 SRGEGSSSLHASSFTPGTSTSVSSLDSDSDSSPSHK-----KLGESKQQRKARHSH 1681
 QY 1156 ---TVSIQDSSEDMIVQNSNESISIQFRTRE-----STFOI----- 1268
 DB 1682 GPLLPTIEDSSBEEL-REEEELLKQEQRELEQOQKSSSSKSKKDKDELRAQRRRR 1740
 QY 1185 -----QDVEVLEPLKCELYSGE-STGNCEDRLPVKGTEANGKPSQOKLEERP 1232
 DB 1741 PKTPPSNLSPIEDASPTTELRAAAEMELHRSCSEYSPSIESDPGFEISPEKIIEVOK 1800
 QY 1233 VNKSDQIKLKNITDKKNENRESEKKGOT-----STFOI----- 1268
 DB 1801 VYKLPVAVSLYSPDTEK-LIGALKESGQKTLKSAEVEEEMHKTHTKSKSFQIASEKDE 1859
 QY 1269 -----NGKDNK-----PKI 1277

```
Db 1860 VFEKESLYGMLIEDYIESLIEDTYNGTVDITNLAMRQDESNEYIQORGKCKIRASEQI 1919      ||: | :|
Qy 1278 YLGECLKEISERVSGNVPEKVNINKIIPENDIKSLTVKESAIRPFINGVIMEDFN 1337      ||: | :|
Db 1920 YDEPKITDLOEDYY---SVPE---LCSTVPOEDI---VSSYIIP-ESHEIIVLDST 1967      ||: | :|
Qy 1338 ERNSSEKSHLLSSDAEGNROSL--ETLPSTKESDSTQTTTPSA--SCPESNSVNOVE 1393      ||: | :|
Db 1968 VTSTTEKQLL---DAAAYEELMKKQRMQLTPGSSPTQPTSDLAPTDMKVSSIGEIA 2023      ||: | :|
Qy 1394 D----MEIETSEVKKVTIS-----SPITSEEE 1415      ||: | :|
Db 2024 DSTSLSTSTSAISDVSSLSIALSIPDVKITQHFTAEEIEDEYLDIYAREIQEIIISHET 2083      ||: | :|
Qy 1416 SNLSNDPIDENGLPINKNENGVESKRKTVITEVTTWTSVAT-----ESTVIKVEK 1468      ||: | :|
Db 2084 SMLTYSVEGGAASILPSDTAS-----LTSSTSSVCTDSSSPIDSATTGYVDT 2132      ||: | :|
Qy 1469 GDKOTVSSSENCAKSTVTTTTTTVTVKLSTP--STGGSVD-----IISVKE 1512      ||: | :|
Db 2133 SDVSKLVDSEDIIAQVPFTSTEEYSEVSMPIESVAGATTKPAIASDMDTVHQAAVCLPE 2192      ||: | :|
Qy 1513 QSKTVVTTVTDSTTTGGTLVMTVSKYESTRDVKVLMKFSRPPKTRSGTALPSYRKF 1572      ||: | :|
Db 2193 TAPSVFTTIVIKPKQYASDTIYDISTAE---KDAARKMK-----STVETGLIKI 2239      ||: | :|
Qy 1573 VTKSTKKSIFPLPNDLLKLARKGGIREVPYFNVA---KPALDIWYPSRPTFGITW 1628      ||: | :|
Db 2240 HHEDSHKELSJ---DMTRINLNGATSEQPLCLVASVKEPASETPAVTTPRV---VSK 2292      ||: | :|
Qy 1629 RYRLQTVKSLAGVSLMLRLLWASLRDMDMAAKVPPG----- 1664      ||: | :|
Db 2293 TSTVSMPPSSAPALTskvFLSRSSLDSPAQSPSPSPPPPPPPPPPLPPILPKPAIYP 2352      ||: | :|
Qy 1665 -----GGSTRFTSTETITTEIKRDVGPYGRFEYCIRKIICPGVPTPRETPT 1717      ||: | :|
Db 2353 KKSQIQAPMATAPVPLVTSVATLESAAV-----LKNHVVPVTKTYTPTPPPV 2402      ||: | :|
Qy 1718 PORKGLRSSAL---RP---KRPETPKQGPVLIETWAAEELEWEIRAFARVERKEK 1769      ||: | :|
Db 2403 PPKPSSIPAGLVFHSRTEVTKPIAKPAPVPL----- 2436      ||: | :|
Qy 1770 AQAVEQAKKLEQOKPTVIATSTSTSTSTSTISPAQKVMVAPIS----- 1816      ||: | :|
Db 2437 PIAVHKPA-----ETQPKPIGLSLTS--SMTNLNVSSAEYKIASPTSPLSPHSNKSPRL 2489      ||: | :|
Qy 1817 --GSVTTCTKMVLTKVCSA-TVTFQONKFNHOTEATWVKOGOSNGVGVQVKVLGII 1873      ||: | :|
Db 2490 TKPSOETVVVITLSEPGTPEAIT-----SOAVTSWLEAPSKQIIPQMPQIPF--- 2539      ||: | :|
Qy 1874 PSSGTGTSQOTFTSPQRT-----ATVTIRPNTSGSGGTTSSNQVITGPQIRPGMT 1923      ||: | :|
Db 2540 -----TSSMKAVEIQSMADQSMYITGALQITPITQTSTFEKVPSSKSEAVTEVAKTAS 2594      ||: | :|
Qy 1924 VIRTPLOOSTLGKAIIRTP-----VMQVP----- 1947      ||: | :|
Db 2595 VVKRPVPSVGLSVTITIPPEIPIISDQPRYRENGRFHPLGDVLDLRTLTKVDIEMRDCS 2654      ||: | :|
Qy 1948 -----GAPQVMTQIIRGQPVSTAVSAPNTVSST-----PGQKSLTSA-----TSTSN 1991      ||: | :|
Db 2655 MDLSAVSMARQRMFTSDTSRGPVSTVQPAIINLSTACVADPSISVITEVAVMFTCTATV 2714      ||: | :|
Qy 1992 QSSASQPPRPQGGVQKLTMAQLTQL-----TOGHGG 2022      ||: | :|
Db 2715 SYSASTDSLVDLGHAMTTPQLTTSKHFEPAYRVTSSQPPFVSRDEVPINLSLGTSAHAV 2774      ||: | :|
Qy 2023 NOGLT--VVIQOGOTTQQLQIIPQGVTVLPGCQQLMQAAMPNGTVQRFLEFTLATAT 2080      ||: | :|
Db 2775 TWAATKPVTVPPVPSVNTGWTDL-----STSQEPMEIGAVD-LSTTKSHRTV 2820      ||: | :|
Qy 2081 TASTTTTTVSTTAAGTGQRQSKLSPQMVHQDKT--LPPAQSSSVGPAKQAPQTAOPSA 2138      ||: | :|
```

RESULT 5

```
PCLO_HUMAN
ID PCLO_HUMAN STANDARD; PRT: 5147 AA.
AC Q9Y6V0; O43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Piccolo protein (Aczonin) (Fragments).
DE PCLO OR ACZ OR KIA00559.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-759 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Killmann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin.";
RL J. Cell Biol. 147:151-162(1999).
RN [2]
RP SEQUENCE OF 552-4404 FROM N.A.
RX Kraemer J., Wollam C., Wohldmann P., McGrane B.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT the complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
RN [4]
```

RP SEQUENCE OF 4405-4439 FROM N.A.
 RC TISSUE-Placenta;
 RA MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5].

RP SEQUENCE OF 4405-5147 FROM N.A.
 RA Kalicki J., Elliott G.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -!- SUBUNIT: Interacts with Rabac1/pral and profilin (By similarity).
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing: Named isoforms=2;
 CC Comment-Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note-No experimental confirmation available;
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).

DR InterPro: IPR000008; C2.
 DR InterPro: IPR001565; Synaptotagmin.
 DR PRINTS; PR00360; C2DOMAIN.
 DR PRINTS; PR00399; SYNAPTOTAGMN.
 DR SMART; SM00239; C2; 2.
 DR PROSITE; PS00499; C2_DOMAIN_1; 1.
 DR PROSITE; PS00004; C2_DOMAIN_2; 2.
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT NON_TER 1 1
 FT DOMAIN 400 465 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT P-A-K-P-Q-P-Q-P-X.
 FT C4-TYPE (POTENTIAL).
 FT C4-TYPE (POTENTIAL).
 FT POLY-PRO.
 FT PDZ.
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT QTKLMGEG (in isoform 2).
 FT S -> SGNGLGIRVGGKEIPGHSGEIGAYIAKILPGSGAE
 FT /FTId=VSP_003923.
 FT K -> KPTDGTKVVSHPITGEIQ (in isoform 2).
 FT /FTId=VSP_003924.
 FT G -> GOVMVVQNAS (in isoform 2).
 FT /FTId=VSP_003925.
 FT TAHKS -> SKRRK (in isoform 2).
 FT /FTId=VSP_003926.
 FT Missing (in isoform 2).
 FT /FTId=VSP_003927.
 FT SQ SEQUENCE 5147 AA; 563537 MW; CD5D84990498CD3C CRC64;
 Query Match 2.9%; Score 421.5; DB 1; Length 5147;
 Best Local Similarity 18.1%; Pred. No. 1.8e-07;
 Matches 530; Conservative 390; Mismatches 940; Indels 1067; Gaps 128;
 QY 275 LECVKPPL--EEVP-----EDEWCEVCVAHKVPGVTDVCAEATQKNKPYIRHPIGY 324
 DB 1076 LKVKETLSMEKIPPMVTVDQKQESKLE-----KDKASALQEKKPLPE----- 1119
 QY 325 DRSRRKYFLNRLRIIEEDTENENKKIWIYYSKVQLAEILCDLDNDYAEALCKLLEEM 384
 DB 1120 -----ERKLPIEEKIRSEKK-----PLLEE- 1141
 QY 385 REEIHRRMDITDLTNKARGSNKSFLLAAANEEILESTRAKKGIDINVKSPPEETKQNE- 443
 DB 1142 -----KKPTPEDKLLPEAKTSAPEE---QKHDL--LKSQVQAEKLEG 1181
 QY 444 -----TENDSKDAEKNREFFED---QSLEKSDDKT-----PDDPPEQK 480
 DB 1182 RVAPKTVQEGKOPQTKMEGLPSGTQSLPKR-DKTKTKTKEQPQPCTAKPQEKEDDK 1240
 QY 481 SEVGDFKSKNGELSESPCAGKAGSGSYRIITRLNPDLSKLSQLAKSQQVAAAHANKL 540
 DB 1241 SOTSSQQPKSPGSLDGTGYSSDGISSSLGEIPLIPTDEK-DILKGLKKDSFSQSSP- 1298
 QY 541 FKEGKEVLVNVNQSGLSRSLSTKKEVIMKGNINNYFKLGOEGKYRVYH-----NOYSTN 593
 DB 1299 -----SSPSDLAKLESTVLSLEAQASTLADEKSEKTKQPHVSPPEQPKDQKQTQ 1348
 QY 594 SFA-----LNKHQREDHDKRRHLAHLFCLTPAGEFKWNGSVHGSK---VLITSTLRLT 644
 DB 1349 SLSETLEITISEEIEKSEBERKDTFKDKSQQDIPSSK-----DHKEKSEFVDDITTRREP 1404
 QY 645 ITQLENN-----IPSSFLHPNWSHRANWIKAV-QMCKSPREFALALALECAVKPVV 696
 DB 1405 YDSVEESSESNPVPQRRRTSVGSSSDSEYKQDSQSGSEEDFIRKQIIEMSADEDA 1464
 QY 697 MLPINREFLGLRLHWTLSIEREEKKVKKKK-----729
 DB 1465 SGSEDEFIRNLQKEISSSTESQKKEETKCKGKITAGKHRLTRKSTSTIDEDAGRHSW 1524
 QY 730 KQEEETMQOATWVKYTFPVYKHQVWKQKGEYRVTYGGWSWISKTHYRVFKPLPNTN 789

Db 1525 HDEDEAFDESPELKY-----RETKSQESELVVTGGG-----LRRF----- 1562
QY 790 VNYRKSLGKGNMDEMSDEKRSRSPKKIKI-----EPDSEKDEVKGS-----DAAGAD 843
Db 1563 -----KTELNSTADKYAESOSSOKYSLYFDEPELEMESLTDSPDRGRGEG 1611
QY 844 QNEMDISKIPEKKDDQYKELLDSDKPKCKEPEMEYDDMDKTESHVN-----CQESSQV 897
Db 1612 SSSILHASSFTPGTSPTSVSLDESDSPSHKKGESKQQRKARHPHGPLLPITIEDSSEE 1671
QY 898 DVNVSEGFHLRYSYKXKTKSSKLDGLLERRIKQFTLEKQKLEKIKLEGIGIKGTST 957
Db 1672 EELREEE-----ELLKEQKQREIQOOR-----KSSS 1699
QY 958 NSKNLSSESPVITKAKEGCOSDSMROE-QSPNANNDQPEDLIQGCQSQSDSVLRMSDPHS 1016
Db 1700 KSK-----KDKDELRAQRRERPKTPPNLSPIED-----ASPEELRQAEME 1744
QY 1017 TTNKLYPKRVLDDVRSRSPETKCPKON-SIENDIE-----BKVSDLASRGOEPTKSKT 1069
Db 1745 ELHR-----SSCSYSPSIESDPEGFEISPEIKIEVQVKYKLPTA--- 1784
QY 1070 KGNDFFIDDKLASADDIGTLCKNKKPLIOESDITVSSKSKSALHSSVPKSTNDRDAP 1129
Db 1785 -----VSLYSPTD-----EOSIMOKEGSOKALKSAEMEYEMMHKTHYKAFP 1827
QY 1130 LSRAMD-----FPGKLGCSSESNTLENSDVTISIQDSEEDMIVONSIESIQEFTREQDV 1187
Db 1828 AANERDEVFE-----KEPLYGGMLIEDIYESLVEDTYNGSDVGSLLTROE--- 1873
QY 1188 EVLEPLKCELVSESTGNCEDRLPVKGTANGKKPSQOKKLEERPNVKNCSQDKLKNWTD 1247
Db 1874 -----EENGFMQO-----KGRE---QKIRLSEQIYEDPMOKITDLOK----- 1907
QY 1248 KKNNEPSEKKQORTSTFOINGKONKPKIYLKGECLKEISESRVSGNVEPKVNINKI 1307
Db 1908 -----EYELSHSVVPOEDIVSSSF-----I 1930
QY 1308 IPEN-----DIKSUTVKESAIRPFINGDVIMEDFNER-----NSSETKSHL-----LSS 1351
Db 1931 IPESHEIVDLGTWVSTFEERKLLDADAAYEELMKRQOMQLTPCSSPTQAPIGEDWTEST 1990
QY 1352 SDAGNVRDSL-----ETLSTKESSTQTTT---PSASCPSNSVNQVEDMEIE---TSEVKK 1404
Db 1991 MDFRMPDASLTSSVLSGASLTDSSTASATLSIPDKVITQHFSEIEIEDEYVTDYTRIEQ 2050
QY 1405 VTS---SPITSEESNLSDNIDENGLPINKNENNGESKRTVITEVTTTSTVAT---ES 1460
Db 2051 IIAHESILITYSESESATSVPSDTP-----SLTSSVSVCTTDSOSPITTLDS 2100
QY 1461 KTVIKVEGDKQTVSSSTENCAKSTV-----TTTTVTVKLSTPSTGGSYD---IIS 1509
Db 2101 ITVTYTEPDMWTKFEDSEISSTYPFGSIIDYPEISVSLDRTAPPDGRASADHIVIS 2160
QY 1510 VKEOSKTVTTVTDSTTTGGTLVTSMVTSKEYSTDKVKLMKFSRPKTRSG---TAL 1566
Db 2161 LSDMASSIIIESVVPKPGPVADTVSTDLIIE-----KDPVKKAK-----KETNGIILLEV 2212
QY 1567 PSYR---KFVTKSTKKS---IFVLPNDLLKLLARKGIREVYFNY-----NAKPALDI 1614
Db 2213 EAYRDKKELEABRTKSLSTVEDHPSPSVIALPMKEQL-STTYFTSGEFTGOKEPASQL 2271
QY 1615 WPPVS-----PRPTFGITWRYRLQTVKSLAGVSLMLRLMLLAWDMDMAAKVPP 1663
Db 2272 ---PSGSPSVSSLPKAPRPF-----RS-----SSL---DISAQPPP 2302
QY 1664 GGGSTRTETSETITTTTELIRKRDVGPYIRFEYCIRKICIPGVPTPKETPPQKGL 1723
Db 2303 -----PPPPPPPPPPPPPP----- 2316
QY 1724 RSSALRKRPRTPKQTG--PVIIETWVAEELWEIRAEERVEKEKAQAEQOAKRL 1781
Db 1781 ----- 2316

Db 2317 -----PPPPPLPPPTSPKPTIL----- 2338
QY 1782 EQOKPTVIAF-----STTSPTSTSTSTISPAQKVM-----YAPISGSVTTG----- 1822
Db 2339 TVASPYTTATPLFDVAVTLETTAVLRSNGLPVTRICTTAPPVPVPPKPSISPLSVLTHRP 2398
QY 1823 --TKMWLTKVGSFAVTVFOONKNFHQTATWVKQGSNSGVVOQVKUGLIPSSGTG-- 1878
Db 2399 EPSKPIAPKPVLPQPT-----TTQKPTDHPKPTGLS 2432
QY 1879 -TSQOIF-----TSFQPRATVTRPNTSGSGCTTNS-----QVITGPIQIRPGMIVRT 1927
Db 2433 LTSMTNLNLTSDYKLPSPSPLSPHSNKSSPRFSKSLTETVWITLPS-EPG-----T 2486
QY 1928 PLOQSTLGAIRIITPVNVQPGAPQVMTQIIRQPV-----STAYSAP----- 1970
Db 2487 P-TDSSASQAITSWPL-----GSPSKDLVSV---EPVSVVPPVTAVEIPISSBQTEFVIG 2538
QY 1971 --NTVSTPGOKSLTSATSTSNIOSSASPPRQOQOVKLTMAQLTQLTQGHGNOGLTV 2028
Db 2539 ALQTFSATP-----VTAFSSQAAPTSVTOFLTTEVSKTEVSATRSTRAPSGLSSISI 2591
QY 2029 VI-----OGOGOTTGLOLIPQGVTLPGP----- 2053
Db 2592 TIPEPLALDNIHLEPKQYKEDKGLQVLGVDIRVTPKVEVKTDKCIDLSASTMDVKR 2651
QY 2054 -----GOOL--MQAAMPNGTVQRPLFTP--LAT-TAT-----TASTTTTTVSTTAAGT 2096
Db 2652 QITANEVYGKQISAVQPSIINLSVTSSIVTPSLATETVTFVCTASASTGTESLVA 2711
QY 2097 GEOROSKL-----SPOMQVHQDKTLPPA-OSSSVGPAKAPQPAQPSARQPTOP- 2146
Db 2712 EHAMTTPLOLTTSKHAEPYRIPSDQVFIAREEAPINLSLGPAPAHAVTLAITKPVTPVP 2771
QY 2147 -----QSPAQPEVQTOPEVQTOVTTVSSHVPSEAQPTHAQSKSPQVAAQSQPSNVQ 2198
Db 2772 VGTNGWTDVTSQGI--TDGEVVDLSTTKSHRVTVMDESTSSVMKLIIEDEKPDVLTAG 2830
QY 2199 QSPV-----RVQSPSQT----- 2215
Db 2831 RRAVCCDVVYKLPFGRSCTAQAPATLTPEDRFGYRDHYYQYDRSGPYGYRGIGMKPMS 2890
QY 2216 -----TPQSLPGQSQVQVQTTTSQPIQPHPSLQ 2245
Db 2891 DTNLAEAGHFFYKSNADFSEGTDAVTLTSGRVTTGEVNDYSSKTTGTPY---ETQ 2946
QY 2246 IPSQOGPQSQVQSQSTQTLSSG--QPLNOVSVSSSRPQLQIQOQPOVIAVPQ----- 2298
Db 2947 VIS-GAGISTPQYSTARMTPPPGYCVGSLRSSNGVVYSSVATPTPTFAITQPGSI 3005
QY 2299 LQOOVQVLSIQ-SQVVAQIOAQOQSGVPQO-----IKLQPIQIQQS 2339
Db 3006 FSTTVRDLSGIHTADAVTSLPAMHHSQPMPSRYFITTCASSETDIAVTGIDISASLQITM 3065
QY 2340 SAVQTHOIONVTVQAASVOEQLQVQOQLRQO-----QOKKKQOQIEIKREHTLQASN 2392
Db 3066 ESUTAETIDSVPLTTAS--EVFPEV--VGDESALLIVPEEDKQOQOLDLER----- 3113
QY 2393 QSEIIQOVVMKHNNAVTEHLKOKKMTPAEREENQRMVNCVMKYITLIDKIDKEEKAQAK 2452
Db 3114 --ELLE-----LEKIKOOR-----FAEELEWERQBIOR 3139
QY 2453 KRKREESVEQKRKONATKLSALLFKHKQLRAE-ILKKRALLDKDLOIE-----VQBELK 2507
Db 3140 FREQEKIMVQKLEELQSMKOHLLFQOEERQAGFMNRQETLAQOQLQEQIQOQLQOQLH 3199
QY 2508 RDLKIKKEKDLML-----AQATA---VA 2528
Db 3200 QOLEEQKIRIQYNYDPDPSGTASQTTTEQAILEGOYAALLEGSOFWATEDATTASAWVA 3259
QY 2529 APCPP-----VTPVLPAAPPAPPSPPPPGVQHTGLIST-----PTLPVASOKRK 2573
Db 3260 IEIPQSGOWTVQSDGVTOYI---APP-----GILSTVSEIPLTDVVVREK 3303

```

QY 2574 REEEKSSSKKKKMMSTTSKTKKDKLYCICKTPYDESKFYIGCDRCQWYHRCVG 2633
D 3304 QPKRSSGAKVRGQ-----YDDMGENTDDP----- 3329
QY 2634 ILOSEAEILDEYVQCOSTEDAMTVLPLTEKDEGLKR-----VLSLQAHKAWPFL 2688
D 3330 --RSPFKKIVDSGV-----QTDD-----EDATDRSVSRRRRTKKSVDTSVQ----- 3368
QY 2689 EPVDPNDAPDYGVKEPMDLATMEERVQR--RYYEKLETFEVDATK 2733
D 3369 --TDEDODEW-----DMPTRSRKARVKYGDMSME--ADTK 3403

RESULT 6
ID BAZA_MOUSE STANDARD; PRT; 1850 AA.
AC Q91VE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-I interacting protein 5) (TFI-I interacting protein
DE 5) (Tip5)
GN BAZ2A OR TIP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423468; PubMed=11532953;
RA Strohn R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,
RA Laengst G., Grummt I.;
RT "NORC-a novel member of mammalian ISWI-containing chromatin remodeling
RT machines";
RL EMBO J. 20:4892-4900(2001).
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI. May serve a specific role in maintaining or
CC altering the chromatin structure of the rDNA locus.
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
CC (nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Co-localizes with the basal RNA
CC polymerase I transcription factor UBF in the nucleolus.
CC -!- SIMILARITY: BELONGS TO THE WAL FAMILY.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ309544; CAC69992.1; -
CC MGD; MGI:2151152; Baz2a.
CC DR GO:0005731; C:nucleolus organizer complex; ISS.
CC DR GO:0003677; F:DNA binding activity; ISS.
CC DR GO:0030528; F:transcription regulator activity; ISS.
CC DR GO:0006338; P:chromatin modeling; ISS.
CC DR GO:0006355; P:regulation of transcription, DNA-dependent; ISS.
CC InterPro; IPR000637; AT_hook.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR004022; DDT_dom.
CC InterPro; IPR001739; Methyl-CpG_bind.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF02178; AT_hook; 4.
CC Pfam; PF00439; bromodomain; 1.

```

```

DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 4.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
KW Nuclear protein; Repeat; DNA-binding.
FT DOMAIN 505 557 MBD.
FT DNA_BIND 602 614 A.T HOOK 1.
FT DNA_BIND 623 635 A.T HOOK 2.
FT DOMAIN 800 862 DDT.
FT DNA_BIND 1137 1149 A.T HOOK 3.
FT DNA_BIND 1351 1363 A.T HOOK 4.
FT ZN_FING 1623 1673 PHD-TYPE.
FT DOMAIN 1755 1825 BROMODOMAIN.
FT DOMAIN 613 738 LYS-RICH.
FT DOMAIN 647 774 COILED COIL (POTENTIAL).
FT DOMAIN 1006 1012 POLY-GLU.
SQ SEQUENCE 1850 AA; 205597 MW; E5DD4FED1D46DAE1 CRC64;

Query Match 2.9%; Score 419.5; DB 1; Length 1850;
Best Local Similarity 18.0%; Pred. No. 6.6e-08;
Matches 386; Conservative 295; Mismatches 776; Indels 693; Gaps 81;

QY 956 STNSSKNLSESPVI-----TKAKEGCSQSDSMRQEQSPNA 989
D 61 SANPNLKDPLLSQFPGGQYPLNGCGLGNRQSPSPSHNTNLRAGSQEFWANGTSPMG 120
QY 990 NNDQPEDLIQCSQSDSSVLNRMDPSHTTNKLYPKDRVLDVDDVIRSPETKCPKQSIEND 1049
D 121 LNFDSQELYDSPDQNFVMPNGPPSPFFTS---PQTSPLMGSSI---QTAFPSQD-VSSD 173
QY 1050 I-----EEKVSLASRGQEPKSKTKGNDFIDDSKLASADDITGLCKKKKPLIOEE 1102
D 174 IHPDFAAEKELTSVVAENG-----TGLVGSLEEEQELKMC 211
QY 1103 SDTISSSSKSLHSSVPKSTNDRDATPLSRAMDFEGKLGCDSESNSTLSESDTVSIQDS 1162
D 212 GYNGSVSSVESLHQEVSVLVPD---PTVSCLD-----DPSHLPDQ 248
QY 1163 SEEDMIVONSNEISSEQFTREQDVEVLEPLKCELVSGESTG-----NCDRLPVKGT 1215
D 249 LEDTPILS-----EDSLEPFDSLAAEPVSGSLYGIDDAELMGAEDKLPLEG- 294
QY 1216 EANGKKPSQOKLEERPYNKCSQDIKLNKNTTOKKNENRESEKKGORTSTFOINGKONKP 1275
D 295 -----NPVISALD----- 302
QY 1276 KIYLAGECILKEISRVVSGNVPEKVNIN---KIIPENDIKSLTVKESAIRPFINGDVIM 1333
D 303 -----C-----PALSANAEFLADDSQTSASIFVSPSPVLGESVL 340
QY 1334 ED--FERNSSSTKSHLLSSDAEGNYRDSLTLPSTKESDSTQ-----TTTTSASCP 1384
D 341 QDNSEGLNSCSDSEQEIEIETQSS--NFORPL-TEPAPQPPSTQLHPAVSPASPAASLT 397
QY 1385 ESNYSNVQEDMEIETSEVKKVTSSPTSEESNLSNDFIDENGLPINKNENYNGESKRKT 1444
D 398 ASAEISPA-----VSPVASSVPV----- 415
QY 1445 VITEVTTMTSTVATESKTVIKVEKGDQKQTVVSSSTENCARKSTVTTTTTKLSTPSTGGS 1504
D 416 --PEVFVAVSPASSPALPAISLE-----ASMTTPVTSQGSPEPSPAAA 457
QY 1505 VDIISVKEQSKTVVTTTVDTSLTGGLTVTSMT---VSKEYSTRDKVKL---MKFSRPK 1558

```

```
Db 458 FQTVSPARKNVSAAPARADRETTGCAVAVSGSDVLRKRIATPEEVELPLQHGWRREV 517
QY 1559 KTRSGT-----ALPSYRKFKVTSTKKSIFVLPNDLKLKARKGIRE 1600
Db 518 RIKKSHRWQGETWYTGPCGKRMKQPPEVIKYLNRNVHVS-----RRE 561
QY 1601 VPFTNNAKPAIDWIPSPRPTFGITWYRLQTVKSLAGVSLMLRLWASLRWDDMAK 1660
Db 562 --HFSF-----SPRMPVGDFEER-DTPEGLQ-----WQLSAEIPSR 597
QY 1661 VPPGGSTRDETETETITTEI--IKRRDVGPGIRFEYCI-----RKIICPIGVPE 1710
Db 598 IQAITKGRPRNERRAKNKEVPKVRGRGPRPKIKMPELLNKTDMRLPKKLETOILSE 657
QY 1711 TPKETPTPKQKGLRUSALRP-----KRPETPKOTGPIVITWVA 1749
Db 658 DDKAKMTKNNKKRQKVRQGESQTPVQOGARKRKQDTKSLKQKDTKKL-----KA 709
QY 1750 EEELEWEIRAFERVEKEKAQAVEQAK-----KRL-EQOKPTVIA 1790
Db 710 EKEKMTKQKLEKKEKREKVKAKGEGPRAPRSCRADKTLATQKRLKEEQOQAIL 769
QY 1791 TSTSPSTSTSTISPAQKVMVAPISGSVYTGKMLVTKVSGPATVTFQOKNKFQTF 1850
Db 770 EEMKPTGEG--MCLSDHQLP-----PDFTRIPGLTSSRAFSDCLTIVEFLHSFGKVLG 821
QY 1851 TWKQGSNGVQV-----VQO-----KVLG-----I 1872
Db 822 FDLTKDVPISGLVQEGILLCOGSLDKVQDLLVRLKAAALHDPGLPYCYOSLKLILGEMSE 881
QY 1873 IPSSTGTSGQTFSTF-----QP-----RTATVTIRPNTSGSGGTT-----SNSQVITG 1915
Db 882 IPLTRDNVSEILRCFLMAYRVEPFCDLSLTQFPQAPPOOKAAILAFLVHELNSITII 941
QY 1916 PQI-----RPGMTVIRTPLOQSTLGKAIIRTPVMVQGAQVMTQIIRG---QP 1962
Db 942 NEIDKTLTSSVSSCRKNKWIVEGLRR-----LKTALAKRTGRPEVMGAEADGLGRRR 994
QY 1963 VSTAVSAPNTSVSTPGOKSLTS-----ATSTSNIOQSSAOPPRPQOGVOKLTMAQTLT 2017
Db 995 SSRIMEETSGLEEEEEEETAVHGRGRKEGIDVAASSIPELERHIEKLSKRLQFFRK 1054
QY 2018 QHGGNGOGLTVIIOGQGTTLQOLIP--QGTVLPGPGQOLMOAMPNGT---VORFLF 2072
Db 1055 KLLHSSQMLRAVSLGQDRYRRHYWVLPYLAGIFVEGSEGTVEIDEIKQETESLMEVYTS 1114
QY 2073 TPLATTATTATTTTSTTTT---AAGTGEORQSK---LSPQMOVHQDKTLPPAOSSSVGA 2127
Db 1115 TPSSARASVKRELTSNASTSPARSGRPRKPKGSLQOQ---HLQSTIRECDS----- 1165
QY 2128 KAQPOTAQPSARPOQTPQSPQAQPEVQTOPEVOTQTTVSSHVPSEAQP-----THAQS 2181
Db 1166 ---EQAGTVHPPEQPOLQAPTOPLQPS-----SGFLEPGSPFSLGQSOHDLS 1212
QY 2182 SKPQVAAQSPQSN-----VQGSQSPVRVQS--PSQTRIRPS--TFSQLSPGQSOQVOT 2230
Db 1213 QSAFLSNLSQTSQSHNLSLSSVLTAPDSPPGKLDSPQSLEPEDEPDEAQCOPGQPFWN 1272
QY 2231 TTSQ-----PIP-----IQPHTSLQIPSGQSPQSPQ-----VQSSQSTLSSGOTL 2271
Db 1273 FSAQIPCDAAPTPPPAVSEDQPTPSLQLLASSKPMWPTGAANPCSPVQLSTHLPGG-TP 1331
QY 2272 NQVSVSPSRPQLQTOQPOQVIAVPOLOQOVQVLSIQSQVVAQIQAGVGPQ----- 2327
Db 1332 KRLSGDSEMSQSPGLQPKRRGR-----PSKFFKQVEHYLTQLTAAQ--IPPENCSG 1385
QY 2328 -----IKLQPIQIQSSAVQTHQIQNVTVQAASVQ-----EQLORV 2365
Db 1386 WMIWDPETDLVLLKALHPRGIRE--KALHKLKSHKDLQFQVCIQPLTDPIFEPNEUPAL 1444
QY 2366 QO--LRDQOQKKQOQIETKREHTLQASNOSEIIQOVVMKINAVIEHLKQKSMTPARE 2424
Db 1445 BEGWSMSPKKEKTYETDL-----AVLQWVEELEQORVVLSLIQINGWTCPTPDS 1492
QY 2425 ENORMIVCNOVMKYITLDRIDKEEQOAAKKRRESVEQKRSQKQATKLSALLFKHKEOLR 2484
Db 1493 TREDITYCEHLP-----DSPEDIPWGRGREGTVPQ--RONNNPLDLAVM----- 1535
QY 2485 AEILKRALDKDKDQIE--VQBELKDKLKKKEKMDQLAQATAVAAPCPVTPVLPAPPA 2543
Db 1536 -----RLAVLEQNVERRYLRPLMAAHEVVVEKALLS----- 1567
QY 2544 PPPPPPPPGVQHTGLLSTPTLPVASO--KRRREBEKSSSKKKKKMISTTSKETKKDT 2601
Db 1568 -TPNGAPDGTSTFEIYEITPRVRVWRQTLERCRSAQVCLCMGQLERSIAW-----EKS 1622
QY 2602 KLYICKTKPYDESKFYIGDCRQNNYHGRCVIGILQSEAEILDE--YVCPQC--QSTEDAM 2657
Db 1623 KVTCLVCRKGDNDDEFLLCDGDRGCHIVC---HRPKMEAVPEGDWFCVCLSQVVEEY 1679
QY 2658 TVLT-----PLTEKDYEGLEK----- 2673
Db 1680 TORPGFPRKGRKKSFFLTFPEGDSRRRMLSRSDSPAVPYRPEDGLSPKRRRHSMRS 1739
QY 2674 -----VLRSLOAKMAMPFLEPDPNDADPYGVYIKPEPMDLATMEERVQRYRYEK 2723
Db 1740 HHSDDLTFCEIILMEMESHDAAMPFLEPNRLVSGYRRVKNPMDFTSMRERLLRGVTS 1799
QY 2724 LTFEVDAMTKIFDNCRYNPNPSDPFYQCAEVLESFFVOKLKGFKASRSHN 2773
Db 1800 SEEFAADALLVFDNCQTFNEDDSEVGKAGHVMMRRFFESWEEFYQOGKQAN 1849
RESULT 7
KI67_HUMAN
ID KI67_HUMAN STANDARD; PRT; 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antigen KI-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schluter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RT J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RL CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
RL CC PROLIFERATION.
CC CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO
CC CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY
CC CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL
CC CC CHROMOSOMES.
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event-Alternative splicing; Named isoforms=2;
CC CC Name=Long;
CC CC IsoId=P46013-1; Sequence=Displayed;
CC CC Name=Short;
CC CC IsoId=P46013-2; Sequence=VSP_004298;
CC CC -!- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS
CC CC PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL
CC CC CYCLE, WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.
```


CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -I- SIMILARITY: TO S.POMBE SPBC215.13.
CC -I- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC

DR EMBL; 238061; CAA86176.1; -
DR EMBL; M16164; AAA35014.1; -
DR EMBL; M16165; AAA35015.1; -
DR EMBL; X13857; CAA32069.1; -
DR PIR; S48478; S48478.
DR SGD; S0001458; MUC1.
DR GO; GO:0003886; C:plasma membrane; IDA.
DR GO; GO:0030447; P:filamentous growth; IDA.
DR GO; GO:0007125; P:invasive growth; IMP.
DR GO; GO:0007124; P:pseudohyphal growth; IMP.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.
FT DOMAIN 210 1367 SER/THR-RICH.
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2BDB61AA9D CRC64;

Query Match 2.8%; Score 401.5; DB 1; Length 1367;
Best Local Similarity 21.2%; Pred. No. 2e-07;
Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;

QY 1135 DREG---KLGCDSE-----SNSTLENSDVTVSIOQSSDIEDMIVQNSNESISEQFRT 1182
DB 191 DFPFGYWNIDCNDNCGGKSSSTTSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSST 243
QY 1183 REQDEVLEPLKCELVSGSTGNC--EDRLPVKGTGANKKPKSQKKLEERPVNCKSDQI 1240
DB 244 TSTSESTSSSTTAPATPTTSCKEKPTPTTSCKEKPT-----PHDTPCT 296
QY 1241 KLNKTTDKNNRESEKKGQRTSTFQINGKDNKPKIYLGKCLKEISESRVVSNGVNEPK 1300
DB 297 KKKTTTSKTKTKTTTPVTPSSSTE-----SSSAPVTPPSSSTTESSAPVTSSTE-- 350
QY 1301 VNNINKIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSSSTKSHLLSSDAEGNVRD 1360
DB 351 --SSSAPV---TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP 394
QY 1361 SLETLPSKESD---TOTTPSACPSNVSNOVEDMEIETSEVKKYTSSTPSE--- 1413
DB 395 VTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV---TSSSTTESSAPVTSSTTES 453
QY 1414 -----EBSN--LSNDFIDENGLPINKNENNGESKRKTVITEVTMTST-VA 1457
DB 454 SSAPVTPPSSSTTESSAPVTSSTTESSAPVTPPSSSTTESSAPVTSSTTESSAPV 513
QY 1458 TSKTVIKVEKDGKQTVSVSTENCASVTVTVTVTVKLSTPSTGGSDIISVKQSKTV 1517
DB 514 TPSSSTTESSAPAPTSSSTTESSAPVTSSTTESSAPVTPPSS-----STTESSSP 568
QY 1518 VTTTVDLSITTTGGTLVTSMTYSKE--YSTRKVKLMKFSRPKKTRSGTALSYKRFVK 1575
DB 569 VTSSTTESSAPVTPPSSSTTESSAPVTPPSSSTTESSAPAPTSSSTTESSAPVTS 628
QY 1576 STKKSTFVLPNDLAKLARKGGIREVPFNYNAKPALDIWPPSPRPFGITWRYRLQPV 1635
DB 629 STTESSA-----PVPTFSS-----TTE 647

QY 1636 KSLAGVSLMLRLWLWASLRWDDMAKVPPGGGGSTRRTTSETETTT-----TEIKR 1685
DB 648 SSSAPVPTP-----SSSTTESSAPVTPPSSST-TESSAPVTSSTTESSAPVTSSTTE 701
QY 1686 RDVGVPVGIREFXCIRKIICIGVP-----EPKETPTQKGLRSSAL----- 1728
DB 702 SSSAPVTPPSSSTTESSAPVTPPSSSTTESSAPVTPPSSSTTESSAPVTSSTTESS 761
QY 1729 -----RPKRPETPKQTGPVITWVAEELELEMBIRAFABRVEKEKAQAQVEQAQKLEQQ 1784
DB 762 APVTPPSSSTTESSAPVTPPSSSTTESSAPVTPPSSSTTESSAPVTPPSSSSNITSS 821
QY 1785 KP--TVIATSTRS-----PTSSSTSTSPQAQVMVADISGVTGTAKMLTKVKGS-- 1833
DB 822 APSSTPFSSSTESSVVPVTPPSSSTTESS-----APVSSSTTESSAPVTPPSSSN 874
QY 1834 -----PATVTFQONKNFHOFATWKKGOGNSGVQVQOKVLG-----IIPS- 1875
DB 875 IYSSAPSSIPFSSSTTESFST-GTIVTPSSSKVPGSQTETSVSSTTTTIVPTKTTTSVTT 933
QY 1876 -----STGTSQ--QTFTSQPRTATVIRPNTSGSGGTTNSQVITGPQIRPGM 1922
DB 934 PSTTTITTTVCSTGINSAGETTSGCSPKVTITV-PTTTTTSVTSSTTIT-----T 985
QY 1923 TVIRPPLQ---OSTLG---KAIIRT--PVMVQPCAPQOVMTQIIRGQPVSTAVS-----AP 1970
DB 986 TVCSTGTSAGETTSGCSPKTTTTPCSTSPSETASESTTSTPTTPVTTVVSTTVVTTTE 1045
QY 1971 NVVSTPPOGKSLTSATSTNSIOSS--ASOPPRPOQGV-KLMAQLTOLTQGHGNOGLT 2027
DB 1046 YSTSKAPGGE-ITTFVTKNIPTTYLTITAPSPSVTVTNFTPTTTTTCVSTGTNSA-- 1102
QY 2028 VVIQGGQTTGLQLIPQGVTVL---PGPQOOLMOAAMPNGTVORFLETPLATTTATTAS 2083
DB 1103 -----GETTSGCS--PKVTITVPCSTGTGEYTEA-----TTLVTTA--- 1138
QY 2084 TTTTTVSTTAAGTGEORQSKLSPQMVHODKTLPPAQSSSVGPAKAPQAPQASARPQO 2143
DB 1139 -VTTTVTTTSTESSTNSAGKTTTG---YTTKSVPTTYVTTLAPS----- 1178
QY 2144 TQPSAPQAEVQTPQEVQJOTTVSSHVPSEAOPTHAOSKPOVAAQSPQSNVQOGSPVR 2203
DB 1179 -----APVTPATNAVPTTITTECSAATNAAGETTSVCSAKTIVSSASAGENTAPSAT--- 1231
QY 2204 VQSPQTRIRPST--PSQLSPGQSQOVQTT---TSQPIPIQPIHTSLQIPSQGQSQOPV 2258
DB 1232 --TPVTAL-PTTVITTESSVGTNSAGETTTGTTTKSIPTTYITTL-IPGNGAKNVEV 1287
QY 2259 QSTQTLLSGQTLNQVSVSSPSRPQLIQOQPQVIAVPQLOQOVQLSQIQSVVQAIO 2318
DB 1288 ATAT-----NPISIKTTS-----QLATTASASSVAPV 1315
QY 2319 AQSGVPPQIKLQLPIQIQSSAVQTHQIENV-VTVQAAS 2357
DB 1316 TSPS-----LTGPLQASGSAVATISVPSISSTYQGA 1348

RESULT 9
BA2B_CHICK STANDARD; PRT: 2130 AA.
ID BA2B_CHICK
AC Q9DEI3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2B (Extracellular matrix protein F22).
DE protein F22).
GN BA2B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]

RP SEQUENCE FROM N.A.
RA Yoon H., Philip N.J.;
RT "Cloning of a new extracellular matrix protein expressed in retina.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May play a role in transcriptional regulation.
CC -1- SIMILARITY: BELONGS TO THE WAL FAMILY.
CC -1- SIMILARITY: Contains 1 bromodomain.
CC -1- SIMILARITY: Contains 1 DDT domain.
CC -1- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF224275; AAC36791.1; -;
CC HSP; Q92831; 1891;
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
KW Nuclear protein; DNA-binding.
FT DOMAIN 693 743 MBD.
FT ZN_FING 1010 1072 DDT.
FT DOMAIN 1895 1945 PHD-TYPE.
FT DOMAIN 2039 2109 BROMODOMAIN.
FT DOMAIN 9 46 SER-RICH.
FT DOMAIN 184 208 SER-RICH.
FT DOMAIN 212 219 POLY-GLU.
FT DOMAIN 544 610 ASP-RICH.
FT DOMAIN 797 984 COILED COIL (POTENTIAL).
FT DOMAIN 882 886 POLY-LYS.
FT DOMAIN 1219 1262 ASP-RICH.
FT DOMAIN 1909 1912 POLY-LEU.
SQ SEQUENCE 2130 AA; 236140 MW; 208C48FB0BA68F70 CRC64;

Query Match 2.8%; Score 396.5; DB 1; Length 2130;
Best Local Similarity 17.9%; Pred. No. 5e-07;
Matches 464; Conservative 345; Mismatches 902; Indels 875; Gaps 104;

QY 419 ESIRAKGDIDNVKSPETEDKNE-----TENDSKDAEKNEEFEDQSLKSDSD 469
Db 170 DKIKDKPRKAVESNSDSGSSDTSSEGISSDSDLEDEDEEDQSAESED 229
QY 470 KTPDDDP---EQGKSEV---GDFKSEKSNSELSGAGKAGSGSTRITRPNPDKLSQ 524
Db 230 ESDSENAEHENKNKVLHSGVKMDKTDQKAHEKSQEKTRHQIPLVSDSQTSSFSQ 289
QY 525 LKSOQVAAA---AHEANKLFE--GKEVLVNSQGEISRL-----STKKEVIMK--- 568
Db 290 QKQPVLSQQLPFIFQSSQAEEKSVNKHITSQISTGLVPNVKPLSLVHQTKEAYLKIIV 349
QY 569 -----GNINNYFKLQGEQKRYRHYHNYQSTNSFALNKHQHREDHDKRRHLAKHFCILTP 620
Db 350 PPDLLKAGNKT-----SEESIPLIS-----DVRSKREQYKQTF---P 385

QY 621 AGEFKWN-----GSVHGSKVLTJSTLTLLTITQLENNIPSSPL-HPNASHRAWI 669
Db 386 AALQKQESSKNLKVIAISLSSKPTSCSPAQKLTSLNNHSPFLTNALLGNHPQNGV 445
QY 670 KAVQMSKPREPALARALAILCAVAVPVMLPIWREFLGHTRLRHMTSIEEREKVKKKEK 729
Db 446 -----IOSVIOEVLALTTKOK 462
QY 730 KOEE-EETMQQATWVKYTPFKVHKQKQGEYRVTGYGGSWISKTHYIRFV-PKLPGN 787
Db 463 SQTKNESVAIASSTPFSPLPVNLSACGKK-----TTG-----NRTLVPSTSPVLPGS 510
QY 788 TNNYRKSLGKTKNNMDENDESDDKRCSPKKIKIEP-----DSKDDVKGSDDAKGAD 843
Db 511 -----GKDKPVSNNAVAVKTHCLPSAKLVLEQFGRGVSDAPSKESKSDNDD 560
QY 844 QNEMDISKITEKDDQDKELLDSDSKPCKEPEMEVDDDMKTESHVNCQESSQDVVNVYS 903
Db 561 DDEDEDEDEDDSDSQ--SESDSNSESDDTGSSEDEDDDKD-----QDESDD---T 611
QY 904 EGFHLRTSYK-KTKSSKLDGLLERRIKQFTLEEKORLEKIKLEGGIKGIGKSTNSSKN 962
Db 612 EG--EKTPLKLTGSS-----IKS-----SSIGPVVAHSTPLN 642
QY 963 LSESPVITRAKEGCGSDSMRQEQSPNANDOPEDLLIQCSDSSSVLRMSDPSHTNK-- 1020
Db 643 L-----QVAKTPSSAPAL--CPETQPAVFLGTTSTPLPSSH 678
QY 1021 --LYPKDRLVDVSIKSP-ETKCPKONSTEN---DIEEKVSDLASRGQEPKTSKTKGNDF 1074
Db 679 CGISKRRVYDERELRVPLEYGHQRETRRNFGRGLOGEVAYFAPCGK----- 726
QY 1075 FIDDSKLASADDI--GTLICKNNKPLIOESDTIVSSSKSALHSVPKSTNDRDAPLSR 1132
Db 727 ----KLQVPEVVKGVQWC-----LLKEP-----EYVPCIR 753
QY 1133 AMDFEKLGCDSNSTLENSSDTVSIQDSSEDMIVQNSNESISBQFTREODV---EV 1189
Db 754 AM--EGRRG--RPNPDRQHSR-----ESMRRRGRPPNVGSTF 791
QY 1190 LEPLKCELVSGESTGNCEDRLPVKGTANGKPKSOQKLEERPVKCSQOIKLNTDTKK 1249
Db 792 LDSTDAKLL-----RKLQAQEIARAAQIKLLRLKQK 824
QY 1250 NNNRSEKKGQRTSTFQINGKDNKPKIYKLGCLKEISESRVSVGNVEKVNINKIIP 1309
Db 825 EQARAAKEAKQQA---IMAAEKKR--QKQIKIMKO-----QEKIKRIQOIRM 869
QY 1310 ENDIKSLTVKESAIRPFINGDVIMEDFNERNSETKSHLLSSSDAEGNYSLETLPSTK 1369
Db 870 EKELRA-----QQILEAKKKKEAANAALL--EAEKRIKE-----K 904
QY 1370 ESDSTQTTTPASCPESNSYNQVEDMEIETSVKKVTSPTSEEESNLSNDFIDENGLP 1429
Db 905 EMRRQQAVALLKQHELEHRHLDMERRRRQHMMLKAMEARKKAEERLQKQERDEKRL- 963
QY 1430 INNENNVGSKRKTIVITEVTTTSTVA--TESKTVIKVEK-----GKQTVVSS 1477
Db 964 -NKRLEQRLELEMAKELKPNEDMCLADQAKALPELPRIGLVLSGTSFSCMLIVQF 1022
QY 1478 TENCASSTVTTTTTTLSTPSTG-----GSVDIISVKEQSTVTTVTVDLSLTTT 1529
Db 1023 LRNFGKVLGPDVNTDVSLSLTQELGLNIGDSRGEVDLLVK-----LVTAACDPLGVT 1077
QY 1530 GGTIVTSMVSYKEY-----STRDKVKLM-----KFSRPKTRSGTALPSYRKF 1572
Db 1078 G---YKAKTILGEHLLNVGINRDNVSEILQIFMEAHGQOTELTESLTKAFQA----- 1127
QY 1573 VTKSTKKSIFVLPNDLLKKLARGGIREV-PYFYNNAKPAIDWIWYPSRPTFGITWYR 1631
Db 1128 HTPACKAAVLAFLVNEIL--ACSKSVSVSEIDKNIDYMSNLRDKW----- 1169
QY 1632 LQTVKSLAGVSLMLRLWLWASLRWDDMAKVPVPGGGSTRTTETSETEITTTTEILKRDVGPY 1691


```

Db 3909 TPPTPTGTGTPTTTTPIITTTTTP-TPPTGTQ-----TPPTTPIITTTTTPPTPTGT 3963
QY 1983 TSATSTSNIOSSASQ-PRPQQGVKLTAQLTQLQGHGNOGLVVIQGGOTGQLQ 2041
Db 3964 QTPTTPTTTTTPPTPTGTGTPTTPIITTTT-----VPTPTPTGTQPTTT 4015
QY 2042 LIPQGVTLPGCQQLMQAMPNGTVQRELFPLATT-----ATTASTTT 2086
Db 4016 PTTTTTTPPT-----TPPTGT-QTPTTPTTTTTPPTTTPPTGTGTPTTTT 4066
QY 2087 TTVSTTAAGTGEOQRQKLSQPMQVHDKTLPPAQSSVGVGPAQAQPTAQSARPPQQTQ 2146
Db 4067 TVTPTPTPTGTGTPT-----TPPTTTPPTTTPPTPTGTGTPTTTTPTTTT 4114
QY 2147 QSPAQEVQTPQVQVQTTSVSHVSEAPQTHAQSKKPQAAQSQSNVQGSQSPVRVQS 2206
Db 4115 VTPTPTGTGTPTTTTPIITTTTTPPTPTGTGT-PTTPTTTTTPPTPTPTGTGT 4172
QY 2207 PSQTRIRPS---TPSQLSQSQSQVQTTTSQPIQPHITSLOIPSGOQO-----SQPV 2258
Db 4173 PTTPTTPTTTTTPPTPTGTGTGTPTTSTAPIAELTTSNPPPTSTPTQTSRSTSSPLT 4232
QY 2259 QSSTQTLSSGQTLNQVSVSPSRP 2282
Db 4233 ESTILLSTUPPAIENTSTAPPSTP 4256

RESULT 11
PCLO_MOUSE STANDARD; PRT; 5038 AA.
AC Q9QYX7; Q9QYX6; Q9QZJ0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-
GN derived HLMN protein).
OS PCLO OR ACZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,
RP TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Killmann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin.";
RL J. Cell Biol. 147:151-162(1999).
RN [2]
RP REVISIONS.
RC TISSUE=Brain;
RA Killmann M.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 4502-4682 FROM N.A.
RC TISSUE=Brain;
RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;
RA Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking.
CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin.
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9QYX7-1; Sequence=Displayed;
CC Name=2;

```

```

CC CC IsoId=Q9QYX7-2; Sequence=VSP_003928, VSP_003929;
CC CC TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
CC CC stomach. Not detected in other tissues analyzed including adrenal
CC CC gland, testis and pancreas.
CC CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC CC phospholipids. Calcium binds with low affinity but with high
CC CC specificity and induces a large conformational change.
CC CC -!- SIMILARITY: Contains 2 C2 domains.
CC CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; Y19185; CAB60731.2; -.
CC CC EMBL; Y19186; CAB60732.2; -.
CC CC EMBL; AF181269; AAD55786.2; -.
CC CC HSSP; P04110; 1A25.
CC CC MGD; MGI:1349390; Pclo.
CC CC GO; GO:0045202; C:synaptic junction; IDA.
CC CC GO; GO:0005509; F:calcium ion binding activity; ISS.
CC CC GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
CC CC GO; GO:0005522; F:profilin binding activity; IDA.
CC CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.
CC CC GO; GO:0016080; P:synaptic vesicle targeting; NAS.
CC CC InterPro; IPR000008; C2.
CC CC InterPro; IPR001478; PDZ.
CC CC Pfam; PF00168; C2; 2.
CC CC Pfam; PF00595; PDZ; 1.
CC CC SMART; SM00239; C2; 2.
CC CC SMART; SM00228; PDZ; 1.
CC CC PROSITE; PS00499; C2_DOMAIN_1; 1.
CC CC PROSITE; PS00004; C2_DOMAIN_2; 2.
CC CC PROSITE; PS50106; PDZ; 1.
CC CC Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
CC CC Repeat; Alternative splicing.
CC CC DOMAIN 371 470
CC CC 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
CC CC P-A-K-P-Q-P-Q-P-X.
CC CC 2N_FING 502 526
CC CC 2N_FING 967 990
CC CC DOMAIN 2305 2329
CC CC DOMAIN 4394 4488
CC CC DOMAIN 4607 4705
CC CC DOMAIN 4922 5012
CC CC VARSPPLIC 4829 4833
CC CC VARSPPLIC 4834 5038
CC CC Missing (in isoform 2).
CC CC /FTid=VSP_003929.
CC CC SEQUENCE 5038 AA; 547600 MW; DADA460CF3B40888 CRC64;
CC CC
CC CC Query Match 2.7%; Score 390.5; DB 1; Length 5038;
CC CC Best Local Similarity 17.8%; Pred. No. 2.1e-06;
CC CC Matches 431; Conservative 330; Mismatches 885; Indels 773; Gaps 91;
CC CC
CC CC 717 EREEKEKVKKKKKQ-----EEETMQQATVWYTFPVKQVWVKQGEYRVYGG 768
CC CC 1120 EEKALPADKKKKPPAAEAAPLEKKRPPD-----DOKLPDPAKPSASEGEKRD----- 1170
CC CC 769 WSWISKTHYRVFPK-----LPGNTNVNVRK----- 794
CC CC 1171 ---LLAHVQ--IPEESPIGVASLACEGQQPDPTRPDLPGATPQTLPKDRKESRDT 1225
CC CC 795 --SLEGT-----KNNNDENWSDSKKCSRPKKK----- 823
CC CC 1226 PQOAEGTAKGEGEPSKDRTEKEEDKSD-TSSQQPKSPQGLSDTGYSSDGISGSLGEIP 1284
CC CC 824 -IEPDSEKDEVKG----SDAAKADQNMDSKI-----TEKKDQD 859
CC CC 1285 SLIPSEKDLKGLKKKSFSSQESSPSPSDLAKLESTVLSLEAQASTLVGEAKERTQP 1344

```



```

Db 3299 TDVVVKEEQPKRRSSGAK 3317
RESULT 12
PCLO_RAT
ID PCLO_RAT STANDARD; PRT; 5085 AA.
AC Q9JKS6; Q9JLT1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN PCLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT bassoon.";
RL Neuron 25:203-214(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
RP ALA-4694.
RX MEDLINE=21181819; PubMed=11285225;
RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT "An unusual C(2)-domain in the active-zone protein piccolo:
RT implications for Ca(2+) regulation of neurotransmitter release.";
RL EMBO J. 20:1605-1619(2001).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin.
CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC junctions.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9JKS6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF138789; AAF07822.2; -
CC EMBL; AF227534; AAF63196.1; -
CC HSSP; P04410; 1A25.
DR GO: 00045202; C:synaptic junction; IDA.
DR GO: 0005509; F:calcium ion binding activity; IDA.
DR GO: 0005544; F:calcium-dependent phospholipid binding acti...; IDA.
DR GO: 0005522; F:profilin binding activity; ISS.
DR GO: 0007010; F:cytoskeleton organization and biogenesis; ISS.
DR GO: 0016080; P:cytoskeletal vesicle targeting; NAS.

```

```

DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
DR PROSITE; PS50106; PDZ; 1.
KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW Repeat; Alternative splicing.
FT DOMAIN 372 491 12 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-P-Q-P-Q-P-X.
FT ZN_FING 523 547 C4-TYPE (POTENTIAL).
FT ZN_FING 1010 1033 C4-TYPE (POTENTIAL).
FT DOMAIN 2351 2362 POLY-PRO.
FT DOMAIN 4442 4536 PDZ.
FT DOMAIN 4653 4732 C2 DOMAIN 1.
FT DOMAIN 4968 5059 C2 DOMAIN 2.
FT VARSPLIC 4876 4880 Missing (in isoform 2).
FT VARSPLIC 4881 5085 /FTID=VSP_003930.
FT VARSPLIC 4881 5085 Missing (in isoform 2).
FT MUTAGEN 4668 4668 D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT BINDING ACTIVITY.
FT MUTAGEN 4674 4674 D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT BINDING ACTIVITY.
FT MUTAGEN 4688 4688 V->S: SMALL INCREASE IN AFFINITY FOR
FT CALCIUM.
FT MUTAGEN 4688 4689 VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT CALCIUM.
FT MUTAGEN 4689 4689 M->S: INCREASED AFFINITY FOR CALCIUM.
FT MUTAGEN 4690 4691 VV->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT CALCIUM.
FT MUTAGEN 4692 4693 QN->AA: MODERATE INCREASE IN AFFINITY FOR
FT CALCIUM.
FT MUTAGEN 4694 4694 A->S: NO EFFECT ON CALCIUM-BINDING
FT ACTIVITY.
SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
Query Match 2.7%; Score 388; DB 1; Length 5085;
Best Local Similarity 17.8%; Pred. No. 2.7e-06;
Matches 438; Conservative 334; Mismatches 832; Indels 860; Gaps 95;
QY 717 EREEKEKVKKKKEKQEEEEE-----TMOQATWVYVTPVVKHVKQKGEYRVYTCGWSWI 772
DB 1163 EEEKAISADKKERKPPAEKPPLEKKPIPVDKLPPAEKPLSSEGEKHE-----I 1214
QY 773 SKTHVVRFPK-----LPNTNNTNRYKS-----LPNTNNTNRYKS-----LE 797
DB 1215 LKAHVQ--IPEEPTGKVAAGKEEQPDSPREALPGATPLTLPKAGEKERAQAQPAE 1272
QY 798 GT-----KNNMDENMDSKRCRSPKKI-----KTEPDS 828
DB 1273 GSSKDGQGERSKEKEDKSDTSSSQPKSPQGLSDTGYSSDGLSGSLGETPSLPSD 1332
QY 829 EKDEVKG---SDAAKAGADONEMDTSKITEKDDQVKELLDSDSD-----KPKC 873
DB 1333 EKDLLGLKKKSFQESSSPSDDLAKL-----ESTVLSTLEAQASTLVGEAKKTPQOK 1388
QY 874 EEPMEVDDDMKTESHVNCQESSQVDVNVVNSEGFHRTSYKKTKSKLGLGLLERRIKQFT 933
DB 1389 ISPEKPDQOKT-----QTASETLDTITSEE-EIKESQEKV-SPKKDS-----EQGF 1434
QY 934 LEEKQRLKIKLEGGI--KGIGKTSTNNSKNLSSEPVITKAKE----GCQSDSMRQEQS- 986
DB 1435 PSRKEHKEKPELVDDLSPRRASVDSVEDSESSENPSVVRKRRRTSIGSSSSDEYKQEDSQ 1494
QY 987 -----PNANNOQPEDLIQ-----GCQSD----- 1005
DB 1495 GSGEEDFIRKQIIEMSADEAGSGDEEFIRSQLKEISGVGESQKREAKGKGKGVAGK 1554

```

```

QY 1006 -----SSVLRMSDPSTHTNKLYPKDORVLDVDSIRSPTKCPKONSIENDIEEKVSDLAS 1059
Db 1555 HRLTRKSTSFDDAGRRHSWDED---DEFDESPELKFRETKSQES--BELV---VA 1606
QY 1060 RGOEPTKSKTKGNDFIDDSKLASADDIGTLCKNKKPLIQEES-----DRIVSSKSA 1113
Db 1607 GGGGLRFFKTIELNSTIADKYSESSEKKTILYFDEEPELEMESLTDSPEDRSRGESSS 1666
QY 1114 LH-----SSVPKSTNDRDATPLSRAMDEGKLGCDSESNSTLENSSDTIVSQDSS 1163
Db 1667 LHASSETPGTSPSTSVSLDSDSDSPSHKKGESKQOKARHRSHGFL-----LPTIEDSS 1721
QY 1164 EEDMIVQNSNESISEOFRTRE-----1184
Db 1722 EEEL-REBELKEQEKQRELEQQOKRSSKSKDKDELRAQRERRPKTPPSNLSPI 1780
QY 1185 QDVEVLEPLKCELVSGE-STGNCEDRLPVKGTEANGKPSQOKKLEERPVKNCSDQIKLK 1243
Db 1781 EDASPTBELRQAAMEELHRSSCSEYSPSTESDPEGFEISPEKIIIEVKVKLPAAVSLY 1840
QY 1244 NTDDKKNNESEKKGORTST---FQINGKDNKPKIYLKGECLKELISESRVVGNEVPK 1300
Db 1841 SPTDEQSVMOKEGVQKALKSAEEMYEMMQPKPHYKAFPAANERDEVEFEKEPLYGM--- 1897
QY 1301 VNNINKIIPENDIKSLTVKESAIRPFINGDVI-----MEDFERNSESTKSHLLSSDAE 1355
Db 1898 -----LIEDIYIESLV--EDTYNGSVGSLLTRQEEONGFMQORGEQKVL-----QE 1944
QY 1356 GNYRDSLETLPSTK-----ESDSTQTTTP-----SASCPESNVNOQVEDMETETSEVK 1404
Db 1945 QIYDDPMOKISDLQKEFYELSHSVVPOEDIVSSSYIIPESHEIIVDLGSMVNSTSEBK 2004
QY 1405 VT-----SPTISE-----EESNLSNDFIDENGL-----1428
Db 2005 LLDADSAYEELMRQOVQVTDGSPVQTTTGDDMAESTLDFRQVQASLTSLSILSGASLT 2064
QY 1429 -----PINKNENVNGSKRTVITEVT-----TMTSTVATESKTVIKVE 1467
Db 2065 DSTSSATLSPDKVITQFSAEELEDEYVTDYREIQDIIAHESLILTYSEPSSESATSV 2124
QY 1468 KGDKQTVVSTENCAKSTVTTTITVTKLSTPSTGGSVDLISVKEQSKTVVTVTVDSLT 1527
Db 2125 PSDTSLSTISSYCTTSSPSTLDSLTVT-EPADVMTKFKDSEESISSTFFPGSII 2183
QY 1528 TTGTLVTSM--TVSKEYSTRCKVKLMKFS-----RPKTRSGTALPSYR 1570
Db 2184 DYPEDISVSLDRTIMPESRTNEDRIVLSFGMAPSVSVESVGTKPERQADTISTDLPISE 2243
QY 1571 KFTVTKSTKSIFLVPNDLKLARKGGIREVPYFNYNNAKPAIDWIYPSPRPFTGITWRY 1630
Db 2244 KDLIKG-----RKETGDGIILEY-----LDAYK-----DK 2268
QY 1631 RLQTVKSLAGVSLMLRLWLASLWDDMAKVPPGGSGSTRTETSETITTEITIKRRDVG 1690
Db 2269 REESEAEELTKISL-----PEPGLAQAPSS-----VTAPOIKEQHVSP 2305
QY 1691 YGIRFEYCIRKIICPIGVPEPP-----KETPTPQKGLRSSAL-----RPKRPE 1735
Db 2306 HVSQK-----ISQOEKPTYRLPSGSLPVSTHPSKSRPFRSSSLDISAQPPPPPP 2356
QY 1736 PKQTGPVIETWVAEELELWELRAFAERKEKAQAQAVEQAQKRLKLEQQKPTVIATSTTS 1795
Db 2357 PPPPSP-----STSSPPPTPLPPATS-----2378
QY 1796 PTSSTSTISPAQKVMVAPISGSVTTGTGKWLVTTKVGSPPATVTFQKNKNFHOTFATWV-- 1853
Db 2379 PKPPT-----YPKKLAVA---ATVTSTT--IVTTHVDALTMVEAAAAARRSNGLPATMCA 2429
QY 1854 -----KQGSNSGVV-----QVOQKVLGIIPSSSTGTSTSOQ- 1882
Db 2430 IAPPPVPPKPSQITGLVFTHRPEAIKPPAIKPAVPIQPVTTQKPTDTCPKPTGLSLTS 2489

```

RESULT 13
BA2B_HUMAN
ID BA2B_HUMAN STANDARD; PRT; 1972 AA.
AC Q9UIF8; Q96EAL; Q96SQ8; Q9P252; Q9Y4N8;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bromodomain adjacent to zinc finger domain 2B (hwalp4).
 GN BAZ2B OR KIAA1476.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Testis;
 RC MEDLINE=20130112; PubMed=10662543;
 RX Jones M.H., Hamana N., Nezu J., Shimane M.;
 RT "A novel family of bromodomain genes.";
 RL Genomics 63:40-45(2000).
 RN [2]
 RN SEQUENCE OF 1-776 FROM N.A. (ISOFORM 1).
 RC TISSUE=Melanoma;
 RA Ansoerge W., Wirkner U., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE OF 1-799 FROM N.A. (ISOFORM 2).
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maita M.A.
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RN SEQUENCE OF 427-1215 FROM N.A. (ISOFORM 3).
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
 RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K.,
 RA Masuho Y., Kanehori K.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE OF 753-1972 FROM N.A. (ISOFORM 1/2/3).
 RX MEDLINE=2027482; PubMed=10819331;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVII.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 7:143-150(2000).
 RN [6]
 RN SEQUENCE OF 1524-1972 FROM N.A. (ISOFORM 1/2/3).
 RC TISSUE=Testis;
 RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May play a role in transcriptional regulation
 interacting with ISWI.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing: Named isoforms-3;
 CC Comment-Experimental confirmation may be lacking for some
 CC isoforms;
 CC

CC Name=1;
 CC IsoId=Q9UIF8-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9UIF8-2; Sequence=VSP_000553;
 CC Note=Inferred from Ref.2;
 CC Name=3;
 CC IsoId=Q9UIF8-3; Sequence=VSP_000554;
 CC Note=Inferred from Ref.3;
 CC -!- TISSUE SPECIFICITY: Expressed at varying levels in several
 CC tissues, whereas a smaller transcript was expressed specifically
 CC in testis
 CC -!- SIMILARITY: BELONGS TO THE WAL FAMILY.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
 CC frameshift in position 731.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB32255; BAA89212.1; -;
 CC EMBL; AL834381; CAD39044.1; ALT_INIT.
 CC EMBL; BC012576; AAH12576.1; ALT_FRAME.
 CC EMBL; AK027612; BAB55231.1; ALT_INIT.
 CC EMBL; AB040909; BAA96000.1; -;
 CC EMBL; AL080173; CAB45759.1; -;
 CC PIR; T12495; T12495.
 CC HSSP; Q2831; 1B91.
 CC Genew; HGNC:963; BAZ2B.
 CC MIM; 605683; -;
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR004022; DDT dom.
 CC InterPro; IPR001739; Methyl-CpG_bind.
 CC InterPro; IPR001965; Znf_PHD.
 CC Pfam; PF00439; bromodomain; 1.
 CC Pfam; PF02791; DDT; 1.
 CC Pfam; PF01429; MBD; 1.
 CC Pfam; PF00628; PHD; 1.
 CC PRINTS; PR00503; BROMODOMAIN.
 CC SMART; SM00297; BROMO; 1.
 CC SMART; SM00571; DDT; 1.
 CC SMART; SM00391; MBD; 1.
 CC SMART; SM00249; PHD; 1.
 CC PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 CC PROSITE; PS50014; BROMODOMAIN_2; 1.
 CC PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 CC PROSITE; PS50016; ZF_PHD_2; 1.
 CC KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
 CC Nuclear protein; DNA-binding; Alternative splicing.
 CC
 CC FT DOMAIN 549 601 MBD.
 CC FT DOMAIN 891 953 DDT.
 CC FT ZN_FING 1735 1785 PHD-TYPE.
 CC FT DOMAIN 1881 1951 BROMODOMAIN.
 CC FT DOMAIN 45 69 SER-RICH.
 CC FT DOMAIN 73 79 POLY-GLU.
 CC FT DOMAIN 399 470 ASP-GLU-RICH.
 CC FT DOMAIN 641 672 ARG-RICH.
 CC FT DOMAIN 687 865 COILED COIL (POTENTIAL).
 CC FT DOMAIN 706 875 LYS-RICH.
 CC FT DOMAIN 1100 1143 ASP-RICH.
 CC FT DOMAIN 1138 1179 COILED COIL (POTENTIAL).
 CC FT VARSPPLIC 437 534 Missing (in isoform 2).
 CC FT VARSPPLIC 593 626 /FTID=VSP_000553.
 CC FT VARSPPLIC Missing (in isoform 3).
 CC FT CONFLICT 137 137 S -> F (IN REF. 1).
 CC

FT	226	226	S -> L (IN REF. 1).
CONFLICT	627	627	E -> E (IN REF. 1).
FT	722	722	G -> K (IN REF. 3).
CONFLICT	790	799	RIKEMRRQ -> QKKKKKKK (IN REF. 3).
FT	1128	1128	K -> Q (IN REF. 1).
CONFLICT	1183	1183	Q -> P (IN REF. 1).
FT	1195	1195	Q -> R (IN REF. 1).
CONFLICT	1453	1453	L -> S (IN REF. 5).
FT	1838	1838	K -> Q (IN REF. 1).
SEQUENCE	1972	AA; 220709 MW; D4F105C927B91569 CRC64;	
Query Match	2.7%;	Score 385; DB 1; Length 1972;	
Best Local Similarity	18.1%;	Pred. No. 1.2e-06;	
Matches 477; Conservative 323; Mismatches 888; Indels 950; Gaps 105;			
QY	400	NKARGSNFLAAANEILIESIRAKG--DIDNVKSPETETKDNKNE---TENDSKDAEKN 454	
DB	17	NQEQSNQP-LDARVDKIDKPRKKAEMSSNSDSCTSDTSSEGISSSDSDDLEED 75	
QY	455	REEFDQSLKSDDKTPDDPEQKSE----VGDFKSEKSNGLSELSEFGAGKAGSGSTR 510	
DB	76	EEB-EDQSTIESEDSDSDESAHQKSNQVLLHGIDPKADQKATEKAQEK----- 127	
QY	511	IITRLRNP-----DSKLSQLKSOQV-----AAAAHEANKLFKKGKVLVYN 551	
DB	128	---RIHOPLASESTHFSQSQKQPVLSQQLPFIQSSQAKES-----VNKHTSVIQ 180	
QY	552	SQGEISRL-----STKKEVIMK-----GNINNYFKLGOEKYRVYHNOYSTN 593	
DB	181	STGLSVNVPKPLSVNQAKETVMKLIIVPSPDVLKAGNKNT-----SBE 223	
QY	594	SFALNKHQREDHDKRRHLAKHFLCTPAGEFFKWSVHGSKVL-TISTLRLTITQ----- 647	
DB	224	SSLSL-----ELSKREQYKQAF---PSQKKQESKSLUKKVIKALSNPKATSSSPAHPK 276	
QY	648	--LENNIPSF-LHPNWSHRANWIKAVQCMCKPREFALALAI-----LECAVK 693	
DB	277	QTLNHNHPNFTNALLGNHOPNGV---IQSVIQEAPLATTKTKMQSKINENIAAASS 332	
QY	694	PVVMPLPIWREFLGHTRLHMTSTIEREKEKVKKKEKQEEETMQQATWVKVTFPVKHV 753	
DB	333	TPFSSPNLSTSGRRTPGNOPTVMPASPIILHSQGEKAVSNVNPVKTHHSHPAKSLV 392	
QY	754	WKQGEYRYVTGYGWSWISKTHYRVFVKPLPGTNNVNYRKSLEGTKNMNMDESDKR 813	
DB	393	---EQFRGT-----DSDIPSKSDSEDEDEDEDED----- 423	
QY	814	KCSRPKKIIPDSEKDEVGSDAAKAGADONEMDISKITEKKDQDVKELLSDSDKPK 873	
DB	424	-----EEDDEDE--SDQSSESNSSESDETEGSEEDDDKDDKDESDDTEGE 470	
QY	874	EEMPEVDDDMKTESHVN-----QOESQVDVNVNVSQ 905	
DB	471	KTSMKLN--KTTSVSKPSMSLTGHSTPRNLHIKAPGAPAAALCESQ-----SPA 520	
QY	906	PHLRTSYKKTKTS-----SKDLGLLRRIKQFTLEK-QRLEKIKLEGG-IKGIGKTSN 958	
DB	521	F-LGTSSTLTSSPHSGTSRRRVTDRELRIPLEYGWQRETRIRNFGRLQGEVAYAP 579	
QY	959	SSKNLSESPVITK--AKEGCOSDMROEQSPNAN-----NDOPEDLIQCSQSDSS 1007	
DB	580	CGKKLRQPEVILKLSRNGIM-DLSRDNFSSAKIRVGDYFARDPGQM-QWCLLKEED 637	
QY	1008	VLRMSDPSHTTNKILPKDRVLDDVYSIRSPETKCPKQNSIENDIEEKVSDIASRGQETKS 1067	
DB	638	VI-----PRIMEGRRRGRP-----NPDQRAREESMRERRKGRP--- 674	
QY	1068	KTGNDFRID--DSKL--ASADDIGTLICKNK--KPLIQESDTIVSSSKSALHSSVPK 1120	
DB	675	-NVGNAEFLDNADAKLRLKQAQIAQAAQIKLLRLKQOEQARVAKAEAKK--QQAIMA 731	
QY	1121	STNDRATPTLRAMDFEGKLCDSSESNTLSSSDTVSIQSDSEDMIVQNSNESIQEF 1180	

FT	DOMAIN	1968	1979	POLY-SER.	
FT	VARSPLIC	2333	2371	Missing (in isoform Short).	
FT				/FTid=VSP_003411.	
FT	CONFLICT	1952	1952	I -> T (IN REF. 2).	
FT	CONFLICT	2090	2090	A -> P (IN REF. 2).	
SQ	SEQUENCE	2453	AA; 270640 MW; 52208B40382F7E6A	CRC64;	
Query Match					
Best Local Similarity 2.7%; Score 382; DB 1; Length 2453;					
Matches 434; Conservative 302; Mismatches 782; Indels 782; Gaps 108;					
Qy	298	HKVPGVDCVAETQKKNPYIRH	----	EPICYDRSRKYWFLNRLRIIEED	----- 343
Db	378	HEISEIIDGLSEQENNEKQROLSVIPPMFMDAQRVVKFINNGLMEDPMKVYKDRQM	437	----	
Qy	344	-----TENENE-----	KKIYVYTKVOLAEILICLDKDYWEALCKILEEMREIHR	390	
Db	438	NVMTDHEKEIFDKQFIQHPKFNFLIASIYERKSVPCVLYY	-----	LTKNENYKALVRR	493
Qy	391	HMDITDLTKAGSNKSFLLAANEEILSIRAKKGIDNVKSPETEEDKDNETENDSKD	450	----	
Db	494	NY-----	CKRRGNQOIARPSQOEKVEKEEDKA	-----	EXTEKEEEKKDEEK 538
Qy	451	AENREFEQSLKSDDDTPDDPQKSEYGVDFKSEKSNSELSPOGAGKASGSTR	510	----	
Db	539	DDK--EDSKETTEKDRTEATA-EEPEE-REQVTPRGKRTAN	-----	SQGRGKR	----- 584
Qy	511	IITLRNPDKLSOLKSOQVAAAAHAANKLFKEGKEVLYVNSOGETSRLSTKK-EVIMKG	569	----	
Db	585	-VTR-----	SMTSEAAAANAANAATEEPPLPPPPPISTEPVTSRWTEEMEVAKKG	638	
Qy	570	NINNYFKLQBGKYRVYHNOYSTNSFALNKHQREDHDKRHLAHLFCLTPAGEFKWGS	629	----	
Db	639	LVEHGRNWAATAKAVGTGKSEAQCKNFYNY	-----	KRRH	----- 672
Qy	630	VHGSKVLITSLRLTITTOLENNPSSFLHPNASHRANWIKAVQMSKCPREFALALALE	689	----	
Db	673	-----	NLONLLQO--HKQKASR	-----	KPRE--ERDVQS 697
Qy	690	CAVKPVVMLPTWREFLGHTLRHMSTJEREKEKVKKKKEQEEETMOQATVWYTFPV	749	----	
Db	698	C-----	ESVAST-----	VSAQEDEDIASNEENEPDSEGAENSDTE-SAPS	739
Qy	750	KHQMVKOGEE--	YRVTVGGVSWISKTHV-----	YRVVPKLPNTNVNYKSL	796
Db	740	PSPVEAAKSSDSENAAASRGNTPEPVAELEATTDPAPCASPSSAVP	-----	TTKPAERESV	795
Qy	797	EG-----	TKNMENDESCKRCS--	RSPKKIKIEPDSKEDEKGSDAAG-ADQ	844
Db	796	EAQVTDASAETAEPMDVDHEECGAGSVDLPPAPTAKSDVPDPMQVPENTASKGEGDA	855	----	
Qy	845	NEMDI--	SKTEKKDQDV-----	KELLDSDSKPKCEE-----	PMEV 879
Db	856	KERDLESTSEKTEARDEDVVAEQIERPEQSDSDSSATCSADSEGVDEGEPEQRQVFMDA	915	----	
Qy	880	DDMKTE-----	SH-----	VNCOESSQDVNVVSEGFHLRTS	911
Db	916	KPSLLTPPGSILISSIPKPNLLDLPLQOHRAAVIPPMVSCPCN-IPIGTPVSGYAL	-----	971	
Qy	912	YKKTSSKLGGLERRIKQFTLEKQRLKLEGGIKGIGKTSNKKN	-----	LSES	966
Db	972	YORHIKAMESALLE-----	EQORQEQVDLECR--	SSTSPCSTSKSPNREWEVLQPA	1022
Qy	967	P--VTTKAKGQSDMRQESP-----	NANNQPEDLIQCSQSDSVLRMSDPS	1015	
Db	1023	PHQVITNPEGVRLPTTRPPPLIPSSKTTVASEKPSFIMGSSISQGPGLYS--S	1080	----	
Qy	1016	HTNKLXPDKRDLVDVSIKSPETKCPKQNSIENDIEKVSOLASRQBPPTSKTKGNDF	1075	----	
Db	1081	H--NQAYPQ-----	EAPKPSVSGISLGL-----	PRQESTKA	----- 1110
Qy	1076	IDDSKLASADDIGTLCKNKXPLI--	QESDITIVSSSKSALHSSVPKSTNDROATPLSR	1132	

Db	1111	-----	APLTVIKOEEFS-----	PRSQNSQBPGLLVR	1136	
Qy	1133	AMD---	FEGKLCGDSSENSTLENSDFTVSIQDSSSEDMIVONSNEIS	SEQFRTREDDVEV	1189	
Db	1137	AOHEGVWRGTAGAVQEGSITRGTASKIS	-----	VET	1168	
Qy	1190	LEPLKCELVSGESTGNCEDRLPVKG--	TEANGKPPSOQKLEERPVPVNCSDQIKLKNTT-	1246		
Db	1169	ISSLRGSIQTGTPA-----	LPOAGIPTALYKGPVSRMP	IESSPEKVREEAASKGHVI	1222	
Qy	1247	-DKNN-----	ENRESEKKGQRT--	STFOINGKDNPKI--	YLKGECLKEISERSRVSG	1295
Db	1223	YEGKSGHLLSYDNINKNAREGTRSPRTAHMSLKRSEAVEAGS	IKOGMSRESVP	APLEG	1282	
Qy	1296	NYEPKVNINKIIPENDIKSLTVKESAIRPFFINGVIM	-----	EDFNE-----	R	1339
Db	1283	LI--	-CRALPRGSPHSDLKERTV-----	LSGSIMOGTPRATAEFEDGLKYPQKIR	1331	
Qy	1340	NSETKSHLLSSDABEGNYRDSLETLPSTKESDSTQTTT	PSASCESNSVNOVEDMEIET	1399		
Db	1332	ESPPIRA--	FEGAITKGPYDGITTI--	KEM-----	GRSHEIPRODILIT	1372
Qy	1400	SEVKV-----	TSSPITSEESNLSNDFIDENGLPINKNENVNGSKRKTIVTEVTMTS	1454		
Db	1373	QESRKTPEVOSTRPII--	EGSIS-----	OGTPI-KFDNNSGQSAIKHNKSLITGPS	1422	
Qy	1455	TVATESKTVIK-----	VEKGDQ-----	TVVSTENCAKSTV-----	1486	
Db	1423	KLPRGMLEIVPENIKVVERGKVEDVKA	GPVRAHRTSVVSSGSPVLRLTHEAPKAQLSP	1482		
Qy	1487	-----	TTTTTTITVKLSTPSTGGSV-----	DIISVKEQS--	KTVVTVTVTDSLTGTTG--	1530
Db	1483	GLYDDSSARPTVSYQNTISRGSPMNRISDVSSSKAS	SHERKSTLTPQRESIPAKSPV	1542		
Qy	1531	---	GTIVT-----	SMTVSKEYSTRDKVKL--	MKFSRPKKTSGTAL-----	1566
Db	1543	PGVDP	IVSHSPDPHHRSSAAGEVYRSHLPTHLDPAMPFHRALDPAAVLLORQLSPTPG	1602		
Qy	1567	-PS-	YKRFVTKSKSIFVLPNDDLLKARKGIREVPFYNKAP-	---	ALDITWYPSPR	1621
Db	1603	YPSQYQIYAMENTROTIL--	NDYITTSQOMVNLR--	PQVTRGLSPREOPLGL-PYPAYR	1656	
Qy	1622	PTFGITWYRLOTVKSLAGVSLMLRLWASLRWDDMAAKVPP	-----	GGGSTRTETSE	1674	
Db	1657	GIDLT--	-----	NMPTTILVPHAGGTSTPPMDRI	1684	
Qy	1675	TEITTEI	IKRRDVGYPYIRFEYCIRKIICPIGVPETPKETPTPQKGLRSSALRKRPE	1734		
Db	1685	TYIPGTQVT	-----	FPPRP-----	YNAASLSPGHP-	1709
Qy	1735	TPKQTGPVLIETWVAEELELWEIRAPAEVEKEKAQAVEQAKKRL	EOOKPTVIATSTT	1794		
Db	1710	-----	THLA-----	AAASAER-EREREKERERERERER	-----	1740
Qy	1795	SPTSSTT	STISPAQKVMVAPISGVTGTTRWLTITKVGSPATVTFQONKNFHOTFATWK	1854		
Db	1741	-----	ERERIAAAPADLYLRPGE-----	QFGRPGSHGYV-----	1770	
Qy	1855	QGOSNSGVQVQKVLGIIPS	-----	STGTSCQFTTSPQRTATVYIRPNTSGSGGTTNS	1910	
Db	1771	--RSPSPSVRTQETILQORPSVFOGTNGTS-	VITPLDP-TAOLRIMPLPSG-----	1817		
Qy	1911	QVITGPOIRGMVIRTPLOOSTLGLKAI	IRTPVMVQGPQOVMTQIIRGQPVSTAVSAP	1970		
Db	1818	---	GPSISOGLPASR---	YNTAADALAL-----	VDAAASAP	1848
Qy	1971	NTVSTPGQSLTSATSTNIQSSASQPPRQOQVKLTMAQLT	QTLTQFHGGHNOGLFWI	2030		
Db	1849	QMDYSKTESKESHEARLEENLRSSAAVSEQQOLEQKNLEVE	-----	KRSVQVCV	1898	
Qy	2031	QOGQOTTGQLOLIPQGVTVLPGPQQQLMQAAMPNGTVQ	RELFPTPLATTATTTTTS	2090		
Db	1899	TSSALPSCGAQ--	PHASVVTSEAGK--	KGPPKPSRYEELTRGKTTITTAANFIDVII	1954	

THIS PAGE BLANK (USPTO)

```

Db      1531 S-----NSDPHD-----IPESKDCVLT-----SEEMFSKDK 1557
Qy      1889 PRTAIVTIRPNTSGSGTTSNQVITGPQIRPCMTVIRTPLOOSTLKGKAIIRTPVMVQPG 1948
Db      1558 TFIVRQSHDEISVSMDSASQMLNEBOLEDMQELVRYQOEHOQOATELLRQAHRMOME 1617
Qy      1949 APQQVMTQIIRQPVSTAVSADNTVSSPGOKSLTSATSTSIQSSASQPPRPOQGVKL 2008
Db      1618 ROREDOEQL---QEBIKRLNRQLAQRSSIDNENLYSERERVLEELKQLSLAGREKL 1674
Qy      2009 TMAQLTQTQGHGNGOGLTVITQGGQTTG-----LQLIPOGV--TVLPFGGOOLMOA 2060
Db      1675 CCELNRNSTQTQNGN-----ENQGEVEEOTFEKELDRKPEDVPPEILSNERYALOKA 1727
Qy      2061 ANPNGTVOFLFTPLATTATTAST-----TTTTVSTTAAGTGEOROSKLSQPMOVH 2111
Db      1728 ---NNRLLKILLEVVKTAAVEETIGRHVGLILDRSSKSQSSASLIWRSEASVKSCH 1784
Qy      2112 QOKTLP-----PAQSSSVGP-----AKAQPQTAPQARPQPTQPOSPAQPE----- 2153
Db      1785 EHRVTVDSEISYSGSDMPRNDINWMSKVTEEGTELSQRLYRSGFAGTEIDPENEELML 1844
Qy      2154 -VOTPEVQTQTVSSHVPSEAOPTHAOSSKPOVAAS-----QPQSNVQOSPVVRVQSP 2207
Db      1845 NISSRLQAAVEKILLEAISSETSSQLEHAKVVTQTELMRESFROKQEAATESLKCOEELRERLH 1904
Qy      2208 SQTIRPSTPSQLSPCQ-----QSOVQTTT-----SOPIPQHTSLOIPS 2248
Db      1905 EESRAREQJAVELSKAEGVIDGYADEKTLFEQIOEKTDIIDRLQELLCASNRLQELEA 1964
Qy      2249 QGQPQSQPOVQOSTOTLSSGQTLNOVSSPSRPPQIQPOQPOVIAVPLOQQOVVLQ 2308
Db      1965 E-----QQQIQEERELLSR---QKEAMKAEAGP---VEQ-----QLLQETELMK 2003
Qy      2309 IQSQVVAQQAQSGVPPQIK-LQLPIQIQSSAVQTHOIQNV-----VTVQAASVQEQLO 2363
Db      2004 EKLEVOCAEKVRDLOKQKALEIDVEQVSRFIEQOEKNTELMDLRQONQALEKQLE 2063
Qy      2364 RVQOLRDOQKKQOQIEIKREHTLQASNQSEIQR-QVVMKHNA-----VIEHLKOKK- 2416
Db      2064 KMRFLDEQAIDREHERDVFQEIQKLEQLKAVVPFQPISEHQTRVEVQLANHLKEKTD 2123
Qy      2417 --SMTPAEREENQRMIVCNQMYIILDKID---KEEQA--AKKKREESVEOK----- 2463
Db      2124 KCELLLSKEQLQORDI---QERNEETEKLFRVRELEQALLVEDRKHFGAVEAKPELSLE 2180
Qy      2464 -----RSQNATKLSALLFKHKEQLRAEILKRALLDK-DLQIEVQ----- 2503
Db      2181 VQLQAEARDAIDRKEKITNLEQL-----EQFRELENKNEEVQOLHMOLEIOKKESTRL 2236
Qy      2504 EELKRLKIKKEDLMOLAQATAVAAPCPVTPVLPAPPAPPPPPPPGVQHTGLLSTP 2563
Db      2237 QELEQENKLFKD-DMEKLGIA-----IKESDAMSTQ 2266
Qy      2564 TLPVASQK-----RKREERKD-----SSSKSKKKMISTSTKETKDKLYCICKTPYDES 2614
Db      2267 DQHVLFCKFAQIIQOEKEVEIDQNEQVTKLOOQKITTDNKVIEEKNELIRDET----- 2321
Qy      2615 KFYIGCDRCRNWTHGRCVGLTQSEAEILIDEYVCPQOSTEDAMTVLTPTEKDYGLKRV 2674
Db      2322 --QIEC-----LMSDQECVK-----RNREEIEQNEVIEKLEQOQELANI 2358
Qy      2675 --LRSLOAHKMAWPFLEPVPNDADPYGYVKEPMDLATMEERVQRRYVEKLETFEVDMT 2732
Db      2359 GQKTSMAHSL-----EADS-----LKHQLDVIAEKLALQOQVETANEENTFMK 2405
Qy      2733 KIFDNCRYNPNPSPFYQCAEVLESFVKLGFKASRSHNNKLOS 2778
Db      2406 NVLKETNF-----KMNQLTQELSLKRERESVEKIQS 2437

```

```

; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171.311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4

Query Match      2.7%; Score 384.5; DB 15; Length 3899;
Best Local Similarity 17.2%; Pred. No. 1e-10;
Matches 483; Conservative 475; Mismatches 1035; Indels 813; Gaps 115;

QY      305 DCVAEI-----QKNKP--YIRHEPTGYDRSRKKYW-----FLNRRLIIE 341
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      113 DCSSEVNGCSFVMRTGKPTNLLREEEFGVDVDSYEQAGDSPHLEMMSESLAGKHETE 172
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      342 EDTENENEKKIWTYSTKVQLAELIDCL-----DKDYWEALCKTL 381
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      173 ELNRELEEMRYTYGTEGLQQLQEPFAIKQDGIITQLTANLQARREKDETMRFELELT 232
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      382 EEMRE-ETH-RHMDITEDLNKARGSNKSPFLAANEI-----LESIRAKKG 426
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      233 EQSKLOIQFOOLQASSETLRNSTHSSYAADLQAQKQILTHQOOLEQDHLLDEYQKKKE 292
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      427 DI-----DNVKSPETEKDKHETENDSKDAEKRE-----EFEDQSLEKDS 467
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      293 DPTHQISFLQEKIK-VYEMEQDK-KVSNKKEIQEKETIEELNTKIIEEKKTKLELKD 350
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      468 DDKTPD-----DDPEQKSEVGDFKSEKSNSELSPEAG--KGASGSTRIITRLRNP 518
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      351 KLTTADKLLGELQEQIVQKNOEIKNMKLELTNSKQKROSSSEIKOLMGTVEELQKRNHK 410
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      519 DSKL-----SOLKSQVAAAAHEA-NKLFEGKEKVLVNVNSQGEISRLSTKKEVI 566
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      411 DSQFETDIVRMEQETORKLEQLRAELDEMYGQIQVQMQLIRQHMAQMEEMKTRHKKE 470
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      567 MKGNNNNYFKLGQEGKYRVYHNOYSTNSFALNK--HQHREDHDKRHLAHKCLTPAGEF 624
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      471 MENAURSTYSNI-----TVNEDQIKLMNVAINELNIKLDPTNSQKEKLKEEL----- 516
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      625 KWNGSVHGSKVLITISTLRLTITQLENNIPSSFLHPNWAHSRANWIKAVQMCSPREFALA 684
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      517 ---GLILEEKCALQOLEDLVEEL-----SFSREQ--IQARQTTAEQ-SKLNEAHS 564
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      685 LAILECAVKPVVMLPIWREFLGHTRLRHMSTS----IEREKEKVKKKKKQBEETMQQA 740
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      565 LSTVBDLKAETVSAESRKELELKHAEVNTYKIKLEMLEKEKNNAVLDRMAESQAEALER 624
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      741 TWVKYTFPVKHOVWKQKE---EYRVYTGCGWNSIKTHVYRPVPKLPNTVNVYRKSLE 797
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      625 LRTQLLFSHEBELSKLDELIEHRIN-----IEKLKDLNGLGIHYKQKQID 668
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      798 GTKNMNDNMND--ESDKRKCRSRPPKTIKIEPDSEKDEVKSDAAGAD-----QNEM 847
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      669 GLQNEMSKQIETMFOFKDNLITKQNQILILBSKLQLQOSLVNSKSEEMTLQINELQKEI 728
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      848 DISKITEKK---BDQYKEL-----LDSDDSKPCKEPMEV 879
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

Db	729	EILRQBEKEKGTLEQEOVEQLQKTELLEKQMKKEKENDLQBFQAQLEAE--NSIILKDEKKTLL	787
Qy	880	DDOMKTESHYNCRESSQDVVNVYSEGPHLTSYKKTKSSKGLDGLLERRIKQFTLEKOR	939
Db	788	EDMLKIHTPYSQEB-----RLIFLDSIKSKSDSVKEKEI-EIILINED	831
Qy	940	LEK--IKLEGIGIKGKTSNKNL-----SESPVITKAKEQCQSDMRQE---Q	985
Db	832	LKQOCIOLNEEIEKORNTFSFAEKNFEVNYOELQEEYACLLKVKDLEDKSKNOELEYKS	891
Qy	986	SPANANDPEDLTQGCQSQSDSSVLRMDPSHTTNKLYPKDRVLDDVISRPEI-----	1038
Db	892	KLKALNEE-----LHLQRINPTTKM--KSVFDEQKTFVAETLEMGEVV	934
Qy	1039	-----KCPKONSIENTEEKVSDLASRQOEPTKSKTKGNDOFFIDDSKLASADDI	1087
Db	935	EKDTTELMKELVTKREKLE--LSQRLSULS-----EOLKOHGEISFLNEEVKSLKQE--	986
Qy	1088	GTLICKNKKPLIOEADTIYSSKSAHSSVPKSTNDRDATPLSRAMDFEGKLCGDSSEN	1147
Db	987	-----KEQVSLRCRELEIILHNRA-----ENVOQCDTQVS	1017
Qy	1148	STLENSDTSVIOSSSEDMIVONSNEISIQEFTREQDVEVLEPLKCELVSESTGNCE	1207
Db	1018	SLL-----DGVVTWTTSRGAEGSVSKVNSFGEEKIMVEDKVSPE---NNTVGEESKQE	1069
Qy	1208	--DRLP--VKGTEANGKPKSQOKKLEERPVPNCKSDQIKLKNITDKKNENRESEKKQ	1262
Db	1070	LILDHLPVSKESLRATQPSENDKLQ-----KELNVLKSPQNDLR-LQMEAQ	1117
Qy	1263	-----TSTFQT-----NGKDNKPKIYLKGBCLBEISVRVSGNVEPKVNNKIIP	1309
Db	1118	ICLSLVYSTHVDQVREYMENKD-----RALCSLKEELIFA--QEEKIRELQK-IH	1165
Qy	1310	ENDIKSUTVKESAIRPPING-----VIMEDNERNSSETKSHLLSSDAEGNYRDSLE	1363
Db	1166	QLELOTKMTOET-----GDEGKPLHLHGKQKAVSEBCSYFLOTCLCSVLGEY---	1213
Qy	1364	TLPSTKESDSTQTTTPSASPESNSVNOVEDMETSETSEVKKVTSPSTSEENLSNDFI	1423
Db	1214	-----YTPALKC-----EVNADKEN-SGDYI	1234
Qy	1424	DENGLPINKN---ENVNGESKRKTIVITEVT-----TMTSTVATESKTVIKVEKGD	1470
Db	1235	SENEDPELQDYRYEVQDFQENMMHTLLNKVTEYNKLLVLQTRLSKIWGQQTDMKLEFGE	1294
Qy	1471	KQTVVSTENCASKSTVTTTTTTVKLSPTSGGSVDIISVKESKSTVVTVTVDLSLTTG	1530
Db	1295	-----ENLPKE-----ETFLSIHSQ-----	1310
Qy	1531	GTLVSTMTVSKSEYSTRDKVKLKMFRPKPTRSGTALPSYRKFTYSTKKSIFVLNDDLLK	1590
Db	1311	-----MTNLEDIDVNIHKSLSLODLEKTLBEOQVELESLSISSLOQ-----LK	1355
Qy	1591	KLARKGGIREVPVFNYNAPKALDIWPYSPRPFTGITWRYRLQTVKSLAGVSLMLRLWA	1650
Db	1356	ETEQ-----NYEAE-----IHCLQ--RELQAVS-----	1376
Qy	1651	SLRWDDMAKVPPGGSTRTETSETETITTEIKRRDVGYPGIRFEYCIRKIICPTGVPE	1710
Db	1377	-----ESTVPP-----SLPVDVSVITESDAQRTMYPGS-----CVKKNL--DGTIE	1415
Qy	1711	TPKETPTPQRGLRSSALRPKRPETPKQTPGVIIETWA--EHEELWEIRAFARVEKEK	1769
Db	1416	FSGEFGVKETNIVKLLKQYQOLBEEVAKVIVSMISAFAPQOETLSRISGGKENTASSK	1475
Qy	1770	-AQAVEQAKRLEQOKPTVIASTTSPTSSTSTISPAQKVMVAPISGVSVTGTKMWLT	1828
Db	1476	QAHAVCOEOHYFNEMK-----LSQDQIGQFTETVDVKFEFKPLSKELGHEKREILL	1530
Qy	1829	TKYGSPTATVTFQONKNPHOTFATWVKVQGSNSGVVQVQKVLGIIPSTSTCSQOTTSFQ	1888

QY	1923	TVIRTPLOQSTGLKAIIRTPVMVQPGAQVMTQIIRGQPVSTAVSAPNTVSSTPGQKSL	1982
Db	3909	TPPTPTGTGTPTTTTPTTTTTP-TPTPTGTQ-----TPPTTPTTTTTPPTPTPTGT	3963
QY	1983	TSATSTNSIQSSAQSP-PRPOGQVKLMAQLTQLTQGHGNGQLTVVIOGQGTGQLQ	2041
Db	3964	QTPPTTPTTTTTPPTPTPTGTGTPTTTTPTTTT-----VTPPTPTGTGTPTTT	4015
QY	2042	LIPQGVTVLPFGQQLMQAAMPNGVQREFLTPLAT	2086
Db	4016	PIITTTTTPPT-----TPTGT-QTPPTTPTTTTTPPTPTPTGTGTPTTTT	4066
QY	2087	TVVSTTAAGTGQRQSKLSPQMVHODKTLPPAQSSVGPAPAKAQPAQPSARPQPTQ	2146
Db	4067	TVVTPPTPTGTGTPT-----TPTTTTTPPTPTPTGTGTPTTTTPTTTT	4114
QY	2147	QSPAQEVOTQPEVOTQTVSSHPVSEAOPTHAQSKKPQVAAQSQPSQSNVQSPVRVQS	2206
Db	4115	VTPPTPTGTGTPTTTTPTTTTTPPTPTPTGTGT-----TPPTTPTTTTTPPTPTTTT	4172
QY	2207	PSQTRIRPS---TPSOLSPGQSQVOTTTTQPIQPHSTSLQIPSGQGPQ-----SQOV	2258
Db	4173	PTTPTTTTTPPTPTGTGTGTPTTHTSTAPIAELTTSNPPESSTPTQTSRSTSSPLT	4232
QY	2259	QSSTQTLSSQOTLNQVSVSSPSRP	2282
Db	4233	ESTTLLSTLPPAIENTSTAPPSTP	4256

RESULT 14

US-10-025-380-1068

; Sequence 1068, Application US/10025380

; Publication No. US20030182191A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Benson, Darin R.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Stoik, John A.

; APPLICANT: Wang, Tongtong

; APPLICANT: Jiang, Yugu

; APPLICANT: Smith, Carole L.

; APPLICANT: King, Gordon E.

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; APPLICANT: Skeiky, Yasir A. W.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Vedwick, Thomas S.

; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.47IC14

; CURRENT APPLICATION NUMBER: US/10/025.380

; CURRENT FILING DATE: 2001-12-19

; NUMBER OF SEQ ID NOS: 1129

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1068

; LENGTH: 5179

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-025-380-1068

QY	1233	TVIRTPLOQSTGLKAIIRTPVMVQPGAQVMTQIIRGQPVSTAVSAPNTVSSTPGQKSL	1982
Db	3909	TPPTPTGTGTPTTTTPTTTTTP-TPTPTGTQ-----TPPTTPTTTTTPPTPTPTGT	3963
QY	1983	TSATSTNSIQSSAQSP-PRPOGQVKLMAQLTQLTQGHGNGQLTVVIOGQGTGQLQ	2041
Db	3964	QTPPTTPTTTTTPPTPTPTGTGTPTTTTPTTTT-----VTPPTPTGTGTPTTT	4015
QY	2042	LIPQGVTVLPFGQQLMQAAMPNGVQREFLTPLAT	2086
Db	4016	PIITTTTTPPT-----TPTGT-QTPPTTPTTTTTPPTPTPTGTGTPTTTT	4066
QY	2087	TVVSTTAAGTGQRQSKLSPQMVHODKTLPPAQSSVGPAPAKAQPAQPSARPQPTQ	2146
Db	4067	TVVTPPTPTGTGTPT-----TPTTTTTPPTPTPTGTGTPTTTTPTTTT	4114
QY	2147	QSPAQEVOTQPEVOTQTVSSHPVSEAOPTHAQSKKPQVAAQSQPSQSNVQSPVRVQS	2206
Db	4115	VTPPTPTGTGTPTTTTPTTTTTPPTPTPTGTGT-----TPPTTPTTTTTPPTPTTTT	4172
QY	2207	PSQTRIRPS---TPSOLSPGQSQVOTTTTQPIQPHSTSLQIPSGQGPQ-----SQOV	2258
Db	4173	PTTPTTTTTPPTPTGTGTGTPTTHTSTAPIAELTTSNPPESSTPTQTSRSTSSPLT	4232
QY	2259	QSSTQTLSSQOTLNQVSVSSPSRP	2282
Db	4233	ESTTLLSTLPPAIENTSTAPPSTP	4256

RESULT 15

US-10-171-311-4

; Sequence 4, Application US/10171311

; Publication No. US20030087270A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

[illegible]

```
Db 787 -----IMTSTASTSKTSVLSVDSVSAKSLSTSESNVSSSTSTSLVN----- 831
QY 911 SYKKTKSKLDGLLER--RIKQFTLEEKORLEKIKLEGKIGKIGKTSNKSNIJSESPV 968
Db 832 ---SQSVSSMSGSVSKSTSLDF-INSNISTEKSE-----SVTSTSDSLRTSTS 878
QY 969 ITKAKEGOSDMROEQSPNANDPEDLIQCCOSDSVLRMSDPSTHTNKLPKDRVL 1028
Db 879 LSDSVSMSTSGSLSKSQSLSTSDASTSQSVSDSTNSI-----STSESLSSEGSTS 932
QY 1029 DQVSIKSPKCPKONSLENDLEEKVSDLASRGOPTKSKTKGNDFFIDDKLAS----- 1083
Db 933 ESISI-----SNSINSVASTSKLESOSTSISLSTS-----DKSMSTSESL 975
QY 1084 ADDIGTLICKNKPLIQEEDTIVSSKSKALSHSVPKSTNDRDPT--LSRAMDFEGKLG 1141
Db 976 SDSTST-----SDSVGSLSVAGSQSVSTSTSDSMSTSEMISDSMTSGSLA 1022
QY 1142 CD-----SESNSTLENSDVTVSIODSSEEDMIVONSNEISSEQFRTREQDVE 1188
Db 1023 ASDSKSMYSVSGMSTSQGSTSESLSDSISTSDSKSLSTLSQSGSTSTSTSSSVR 1082
QY 1189 VLEPLKCELVSGESTGNCEDRLPVKGTENGKPKSQKKLEERPYNKCSDOIKNMTDK 1248
Db 1083 MSE---SQSTSGMSTSQSDSTSI-----TSFSD-----STDS 1114
QY 1249 KNNENREBEKKQORTSTFOINGKDNKPKIYLKGECLKEISERVVGNVEPKVNNIKII 1308
Db 1115 KSASTASSEISQSVST-----STSGSVSTSLSTSNERTS 1152
QY 1309 PE-NDIKSLTVKES-----AIRPFG-----DVIMDEFERNSETK-----S 1346
Db 1153 TMSDSTSLSTSESDSTSDSTSDSISAIISGSESTSIJSESNSTSESKSASFLS 1212
QY 1347 HLLSSDAE-----GNYRDSLETLPSTKESDSTQTTTPSASCPESNVQVDEMEIT 1399
Db 1213 ESLSESTSESTSESLSGSTSDSTSLSDSNSESGSTSTSL-SNSTSGSASISTSGAST 1271
QY 1400 SEYKVTSSPITSEESNLSNDFIDENGLPINKNENVE-----SKRT 1444
Db 1272 STYK---SESVSTSLSTSTSLSDSTSLSTSLSDSTSGSKNSLASMSTSDSISTRS 1328
QY 1445 VITEVTMTSTVATESK---TVIKVEKDGK-----QTVVSSSTENCAKSTVT- 1487
Db 1329 ESLSASTSLSGSTSESESGSTSSSESKSDSTSMLSMQSTSGSTSVSTSESLSDSTST 1388
QY 1488 -----TTTTVTYKLTSTPSTGGVDIISVKEQSKVVVTITVDS 1525
Db 1389 LSLASMNQSGVDSNASASQASSTSTSTSESDSQSTSYTSQSTSQSESTSTSLSDS 1448
QY 1526 L-----TTTGGTLVMTSVKEYSTRDKVKLMFRPKPTRSGTALPSYKFKVTKTKKS 1580
Db 1449 TSISKSTSQSGSTSTGASLGSSESDSQSISTSTSEKSESTSLSDSTSTSNSGSAS 1508
QY 1581 IFVLPNDLKLARKGIRVFPYFNNAKPALDIWPPSPRPTFGITWRYRLOTVKSLAG 1640
Db 1509 TSTL-----LNSASASESDSSSTSLSDSTSGMSQSESDS-----QSTSTSLN 1553
QY 1641 VSLMLRLWLSLWDDMAKVPGGGSTRFTETSETITTEILIKRVDGVPYGRFYECIR 1700
Db 1554 SQSTSTIRMTSTIASESVSESTSESGSTSESTSTSL----- 1595
QY 1701 KIICPIGVPTPKPTPPQKGLRSSALRPKRPETPKQTGPVFIETWAVEEELMEIRA 1760
Db 1596 -----SDSQSTSRSTSGASGASTSTSDSRSTSTSTSTSTSTSTSTSTSTST 1637
QY 1761 FAERVEKEQAQVEQAKKLEQOKPVIATSTSTSTSTSTSTSTSTSTSTSTSTSTST 1820
Db 1638 -----SQSMSTSTSTSTSVSDSTSLSDSTSDSTSTSTSTSTSTSTSTSTST 1688
QY 1821 TGTKM--VLITKVGSPATVTFQONKNFHQTATWVKOGQNSGVVQVQKVLGIIPTSTG 1878
```

RESULT 12

US-09-922-217-1068

; Sequence 1068, Application US/09922217

; Patent No. US20020076414A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Benson, Darin R.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Stolk, John A.

; APPLICANT: Wang, Tongtong

; APPLICANT: Jiang, Tuolu

; APPLICANT: Smith, Carole Lynn

; APPLICANT: King, Gordon E.

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; FILE REFERENCE: 210121.471C13

; CURRENT APPLICATION NUMBER: US/09/922, 217

; CURRENT FILING DATE: 2001-08-03

Query Match		2.8%; Score 401.5; DB 10; Length 1367;
Best Local Similarity		21.2%; Pred. No. 3.4e-12;
Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;		
QY	1135	DPEG---KLGCDSE-----SNSTLENSDVTSVSIQDSSEEDMIVQNSNESISEQFRT 1182
DB	191	DPCFYWNIDCNCNGGTKSSTTSSTSESSSTTSSTSESS-----TTTSSSTSESST 243
QY	1183	REQDVEVLEPLKCELVSGESTGNC--EORLPVKTEANGKKPKSOOKKLEERPVKNCSDQI 1240
DB	244	TSSTSESSSTTAPATPTTTSCTKEKPTPTTTSCTKEKPTP-----PHDHTTPTCT 296
QY	1241	KLKNTTDKKNENRESEKKGORTSTFQINGDKNPKIYLKCELKEISESVVSGNVEPK 1300
DB	297	KKKTTTCKTKKTTTPVPTSSSTTE-----SSGAPVPTPSSSTTESSAPVTSSTTE-- 350
QY	1301	VNNKIIIPENDIKSLTVKESAIRPFINGDVIMEDFERNESSEKSHLLSSDAEGNYRD 1360
DB	351	---SSGAPVP---TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP 394
QY	1361	SLELPSRKESDS---TQTTTPSACPSNSVNOVEDMEITSEVKKVTSSPTISE----- 1413
DB	395	VPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV-TSSTTESSAPVTSSTES 453
QY	1414	-----EESN--LNDNDFIDENGLPINKNNVNGESKRKVTITEVTTTST-VA 1457
DB	454	SSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVP 513
QY	1458	TESKTVIKVEGDKQTVSSSTENCASVTVTVTVTVTVKLSTPSTGSGVDIISVREQSKTV 1517
DB	514	TPSSSTTESSAPAPTPSSSTTESSAPVTSSTTESSAPVPTPSS-----STTESSSTP 568
QY	1518	VTTVTVDLSLTGGTLVTSMTVSKB--YSTKDKVKLMKFSRPKTRSGTALPSYKRVFK 1575
DB	569	VTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPAPTPPSSSTTESSAPVTS 628
QY	1576	STKKSIFVLNDLKLARKGIREVPVFNNAKPALDIWYPSPRPFTFGITWRYRLQTV 1635
DB	629	STTESSA-----PVPPTSS-----TTE 647
QY	1636	KSLAGVSLMLRLMLASLRWDMMAKVPVGGSGTRTETSEBITT-----TEIKR 1685
DB	648	SSSAPVPTP-----SSSTTESSAPVPTPSSST-TESSAPVPTSTTESSAPVTSSTTE 701
QY	1686	RDVGPYGRFYCYRKIKICPGVP-----ETPKEPTPQRKGLRSSAL----- 1728
DB	702	SSSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESS 761
QY	1729	----RPKRPEPKOTGPVIIETWVAEELELWEIRAPAEVKEKAQAEQOAKRLBQQ 1784
DB	762	APVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSNTISS 821
QY	1785	KP--TVIATSTS-----PTSSTTISPAAQVMVAPISGVSVTGTGKMLTKVGS-- 1833
DB	822	APSSTPSSSTESSSVPTPSSSTTESS-----APVSSSTTESSAPVPTPSSSN 874
QY	1834	-----PATVTOQKNFHQTATVWKQGSNGVVQOQKVLG-----IIPS----- 1875
DB	875	ITSAPSSIPSSSTTESFST-GTTVTPSSSKYPSQTSVSTTETTVPTKTITSVTT 933
QY	1876	-----STGTSQ--QFTSFQPRATVTRPNTSGSGTTSNQSIVITGQIRPGM 1922
DB	934	PSTTTTITTCVSTGTSNAGETSGCSPKTVTTTV-PTTTTTSVTSTTTT-----T 985
QY	1923	TVIRTPLO--QSTLIG----KAIIRT-PVMVOPGAQOQVMTQIIRQGPVSTAVS-----AP 1970
DB	986	TVCSTGTSNAGETSGCSPKTIITVPCSTSPSETASESTTSTPTPTVTVSVTVTTE 1045
QY	1971	NTVSSTPQKSLTSATSTNTQSS--ASQPPRPQOGV-KLTMALQTLQTLQHGNGOGLT 2027
DB	1046	YSTSTKPGGE-ITTFVTKNIPTLYLTIAPTPSVTVTNTFTPTTITTVTVCSTGTSNA-- 1102
QY	2028	VVIQOGQTTQQLQLIQGVTVL----PGPQOQLMQAAMPNGTVQRFELTFLATATATAS 2083
DB	1103	-----GETTSGCS--PKVITTTVPVPCSTGTGEYTEA-----TTLVTFTA----- 1138
QY	2084	TTTTTSTAAGTGEQROSKLSQMVHQDHTLPPAQSSSVGPKAKAQOQTAQPSARPQPQ 2143
DB	1139	VTITVTTESTGTNSAGKTTTG---YTTKSVPTTYVTTLAPS----- 1178
QY	2144	TOPSPPAQPEVOTQEVOTQTVSSHVPSEAPTHAQSSKPOVAAQSQPQSNVQSQSPVR 2203
DB	1179	----APVTPATNAVPTTITTTTCSAATNAAGETTSVCSAKTIVSSASAGENTAPSAT--- 1231
QY	2204	VQSPSQTRIREST--PSOLSPCQSOVQTT---TSQPIPIOPHTSLQIPSOQSQSQPOV 2258
DB	1232	--TPVTAL-PTTIVTTESVGTNSAGETTTGYTKTSIPTTYITL-IPGNGAKNYETV 1287
QY	2259	QSSTQTLSSGGTQLNQSVPSSPRPQLQIQPQPOVIAVPOLQOQVQVLSQISQSVVAAIQ 2318
DB	1288	ATAT-----NPISIKTFS-----QLATTASASSVAPVV 1315
QY	2319	AQSGVPOQVQIKQLPIQIQOQSSAVOTHQIQNV-VTVQAAS 2357
DB	1316	TSPS-----LTGPLQASGASAVATSVPSISSTYQGA 1348
RESULT 11		
US-09-815-242-12713		
; Sequence 12713, Application US/09815242		
; Patent No. US20020061569A1		
; GENERAL INFORMATION:		
; APPLICANT: Haselbeck, Robert		
; APPLICANT: Ohlsen, Karl L.		
; APPLICANT: Zyskind, Judith W.		
; APPLICANT: Wall, Daniel		
; APPLICANT: Trawick, John D.		
; APPLICANT: Carr, Grant J.		
; APPLICANT: Yamamoto, Robert T.		
; APPLICANT: Xu, H. Howard		
; TITLE OF INVENTION: Identification of Essential Genes In		
; FILE REFERENCE: Prokaryotes		
; FILE REFERENCE: ELITRA.011A		
; CURRENT APPLICATION NUMBER: US/09/815,242		
; CURRENT FILING DATE: 2001-03-21		
; PRIOR APPLICATION NUMBER: 60/191,078		
; PRIOR FILING DATE: 2000-03-21		
; PRIOR APPLICATION NUMBER: 60/206,848		
; PRIOR FILING DATE: 2000-05-23		
; PRIOR APPLICATION NUMBER: 60/207,727		
; PRIOR FILING DATE: 2000-05-26		
; PRIOR APPLICATION NUMBER: 60/242,578		
; PRIOR FILING DATE: 2000-10-23		
; PRIOR APPLICATION NUMBER: 60/253,625		
; PRIOR FILING DATE: 2000-11-27		
; PRIOR APPLICATION NUMBER: 60/257,931		
; PRIOR FILING DATE: 2000-12-22		
; PRIOR APPLICATION NUMBER: 60/269,308		
; PRIOR FILING DATE: 2001-02-16		
; NUMBER OF SEQ ID NOS: 14110		
; SOFTWARE: FastSeq for Windows Version 4.0		
; SEQ ID NO 12713		
; LENGTH: 2344		
; TYPE: PRF		
; ORGANISM: Staphylococcus aureus		
US-09-815-242-12713		
Query Match		2.8%; Score 401; DB 9; Length 2344;
Best Local Similarity		18.0%; Pred. No. 7.3e-12;
Matches 336; Conservative 271; Mismatches 836; Indels 424; Gaps 57;		
QY	797	ECTKNNMDENDESKRCSRPKIKYEPDSEKDEKVGSDAAGADQNMDSKITEKK 856
DB	747	EVTRNSMDSYSTSGTSQSQSVSTSKADSQSASTSGS----- 786
QY	857	DQVKELLDSDSRKPEEPNEVDMDKTESHVNCQESSQV-----DVNVNSEGFHLRT 910

Db 547 LKIIIEQPOPQSGQESGSEIHVEVKAQSLVTSPPAPSPRKTVPASDORRRSCKTAPASS 606
QY 489 EKXNGELSESPGAGKAGSSTRITILRNPDSS-----KLSQLKSOVAAAAHEANKLFK 542
Db 607 SKSQTEVPK-----RGERVATCLQKRVSLRSQHDILQIMCSKRRSGASEAN----- 654
QY 543 EGREVLVANSQGISLSTKK---EVIMKG---NINNYFKLGOEGKYRV---YHNOYSTN- 593
Db 655 ----LIVAKSWADVVKLAGAKQOTVKIKHGPQBSMKNRORRATPKKPGVEHVSQFSTGH 710
QY 594 ----SFALNKHOHREDHDKRR--HLAHKFCLTIPAGEFK-----WNGSVHGSKVILT 637
Db 711 ANSPCTIIIGKAHTERVHVPARPYRVLLNFISNQKMDFKEDLSGIAEMEKTPVKEQPOLT 770
QY 638 ISFLRLTIQLENNIPISSF-----LHPNASHRANWIKAVQCMCK- PREFALALAI 687
Db 771 -STCHAIANSNLLKQOGTDSGEEPLLTSESEFGNVFTSAQNAAKOPSD----- 822
QY 688 LECAVPVVMPLPIWREFLIGHT-----RLHRMTSIERE----- 719
Db 823 -KCSASP-----PLRQCIRENGNVAKTPRNTYKMTSLETKTSDTETEPSKTVSTVNRSGR 877
QY 720 --EKEVKKK--KEKKQEEBETM-----QOATWVKYTFPVKHQVWQKQGEYRVGY 766
Db 878 STEFRNIQKLPVESKSEETNTEIVECILKRGQKATLQO-----RREGEMKEI--- 925
QY 767 GGSWSLKTHTVYRFVPLGNTNVNYSLEGTCKNNMDENDESCK-----RKCS--RSP 819
Db 926 -----ERPFET---YKENIELKEN--DEKMKAMKRSRTWGQKCAPMSDL 964
QY 820 KIKIIPBDS--KDEYKVG--SDAAKAGADONEMDISKITEKKDQDVKELDSDSKP 871
Db 965 TDLKSLPDTLMDKTARGONLLQTDHAKAPKSEKGIKTM-----P 1006
QY 872 CKE-EPMEVDDDDMKTSHVNCQSSQDVVVNVE-----GFHLRTS----- 911
Db 1007 CQSLQP-----EPINTPTHTKQKLKASLGVKGVKEELLAVGKFTRTSGETHTHREPADG 1062
QY 912 ----YKKTK-----SSKLDGLLERRIKQFTLEEKORLEKLEGIGIGIKTNSSK 961
Db 1063 KSRTFKESQKQILDPAARVTGM--KKWPTPEEAQSLDLE---AGFKELFOTPGPSEE 1117
QY 962 NLSESPVITKAKGCGSDSMRQBSQSPNANDQPE-DLIOGCSQSDSVLRMSDPSHTTNK 1020
Db 1118 SMTDEKT---TKLACKSPPESDVPTSTKQWPKRSRLKADVEEFALRLKLTGPSAGAM 1174
QY 1021 LYPKDRVLDVSI-----RSPETKCPKQNSIEN-----DIE 1051
Db 1175 LTPKPAAGDEKDIKAFMGTPVQKLDLAGTLPGSKRQLOTTPKEKAQALEDLAGFKELFQTP 1234
QY 1052 EKVSDDLASRGO-----EPTKSKTKGNDFFIDDSKLASADDIGTLI-CNNK 1095
Db 1235 GHTEELVAAGTKTKIPCDSPQSDPVDPTPTSTKOR-----PKRSIRKADVEGELLACRNL 1288
QY 1096 KPLIQEESDITVSSKSALSHSVPKSTNDRD-----ATPLSRAMDFEGKLGCDSESNTL 1150
Db 1289 MP-----SAGKAMHTPRPSVGEEDIIIFVGTVPQKL-----DLT 1323
QY 1151 ENSSDVTVISQDSSEEDMIVQNSNESIEQPRTRREQDVEVLEPLKCELVSGESTG-NCEDR 1209
Db 1324 ENLTGSKRRPQTPKEEAQALEDLTGFKELFOTPGHTEEA-----VAAGTKTKMPCES 1376
QY 1210 LPVKGTEANGKPSQOKLEERPVNCKSDOIKLKNITDKNNENRESEKKQORTSFOIN 1369
Db 1377 PPESADPTTTRQKPTPLEKRDVQKELSALK-----KLTOTSGETHHTDKVP 1424
QY 1270 GKDNKPKIYLKGECLKEISESRVSGNV-EPKYVNNINKIIPENDIKSLTVKESAIRPFIN 1328
Db 1425 GGEDKSNAFRETAQKLDPAASVGTGSKRHPTK--EKAQPLEDLAGW--KELFQTP--- 1477
QY 1329 GDVIMEDFNERNSETKSHLLSSDABGNVRSLETLPSTKESDSTOTTTTSPACSPSNS 1388
Db 1478 ---VCTDKPTTHEKTTKIACRSQPD-----PVDPTSSSKPQSKR 1513
QY 1389 VNOVEDMETETSEVKKVTSS-----PITSEEBNSLNSDFIDENGLPINK---NENW 1436
Db 1514 SLRKVDVEEFALRKRTSPSAGAMHTPKPVSQEKNIYA--FM---GTPVQKLDLTENL 1568
QY 1437 NGESKRKQITVITEVNTMTSTA-----TESKTVIKVEKGDKQVTVSTENCASK 1484
Db 1569 TGSKRRLQTPKEKAQALEDLAGFKELFOTRGHTEESMT---NDKTAKVACKSQSPOLDKN 1625
QY 1485 TVTTTTVTTKLSTPSTGSGVDIISVKEQSKTVVTTTVDLSLTTTGGTLVTSMTVSKESYS 1544
Db 1626 PASSKRRLTSLG--KVGKRELLAVGKLTQTSGETHTHTTEPDGKSKMAFMEPKQI 1683
QY 1545 TRDKVKLMFSPKTKRSGTA-----LPSYRK--FVTKSTKKSIFVLPND 1588
Db 1684 LOSAASLTGSKRQLRTPKSKSEVPEDLAGFIELFOTPSHTKESMTNEKTTKVSYRASQPD 1743
QY 1589 L-----KKLARKGGIREVPYFNYNAPALDIWPPSPRPRTFGTWTWYRLQTVKS 1637
Db 1744 LVDTPTSSKPOPKRSLRKADTEE-----EFLAFRKQTPSAG---KAMHTPKP 1787
QY 1638 LAGVSLMLR--LLWASLRWDDMAKVPPGGSGTSTETSETETITTEILIKREDVGPYGR-- 1694
Db 1788 AVGEEKDINTFLGTPVQKLDQPGNLP--GSNRLQTRKKAQALELT-----GFREL 1838
QY 1695 -----FEYCIRKIIC--PIGVP-ETPKETPTPQKGLRSS-----ALRPKRP 1733
Db 1839 FQTPCTDNTTDEKTTKILCKSPQSDPADTPTNTKQPKRSLKADVEEFALFRKLTP 1898
QY 1734 -----ETPKQGTGPVIETWVAEELELWEIRAF---AERVE-----KE 1768
Db 1899 SAGKAMHTPKAA-----VGEEK---DINTFVGTPEKLDLLGNLPGSKRRPQTPKE 1946
QY 1769 KQAAVEQAKKLEQOKPTVIATSTSTSTSTISPAQKVMVAPISGVTGTGKWLIT 1828
Db 1947 KAKALEDLAGFKELFOTPGHTEESMTDDKITEVCSKSPQDPVKTP-----TSSQORLK 2000
QY 1829 TKVGSPTVTFQONKFNHOTFATWVKGOSNGSVQVQOKVL--GIIPSTGTSTQSTFTS 1886
Db 2001 ISLGK-----VGKEEVLVPGKLTOTSGKTTQT--- 2028
QY 1887 FQPTATVITRPNNTSSGGT---TNSQVI-----TGPOIRPGMTVIRTP-----L 1929
Db 2029 -----HRETAGDGKSIKAFKSAKOMLDPANYGTGMRWP-----RTPEBAQSL 2073
QY 1930 QOSTLGLKAIIRTPVMVPGAPQOVMTQIIRGQPVSTAVSAPNTVSVSTPGOKSLTSATSTS 1989
Db 2074 EDLAGFKELFOTPDHTEESTTDDKTKI-----ACKSPPPESMDTPTST- 2117
QY 1990 NIOSSASQPPRQOQGVKLT--MAOLTQLTQGHGNOGLTVVIOGOGQTTGQLQIP--- 2044
Db 2118 -----RRRPKTPPLGRDIVEELSALKQLTQ-----THTDKVPGE 2153
QY 2045 -QGVTVLPGCQOLM-----QAAMPNGTVQ-----RFLF-TPLATTATTAS 2083
Db 2154 DKGINVFRETAQKLDPAASVGTGSKRQPTPKGAQPLEDLAGLKGELFQPICTDKRPTH 2213
QY 2084 TTTTTVSTTA-----AGTGE--OROSKLS--POMQVHODKTLPPAQS SVSGPAKAQOTA- 2134
Db 2214 EKTTKIACRSQPDQDPTPIFRQSKRSLRKADVEEESIALKRRTPSVGKAMDTPKAPAG 2273
QY 2135 -----QPSARPOQTPQSPAQPEVQTPQEVQTTQTVSS-----HVP 2171
Db 2274 GDEKMKAFMGTPVQKLDLPGNLPGSKRWQTPKE-----KAQALEDLAGFKELFOTP 2326
QY 2172 SEAQPHQAQSKQVAAQSOPOQSVNQQQSPVVRVQSPQTRIRPS----- 2215
Db 2327 GTDKPT-TDEKTTKIAKCS-PQ-----PDPVDTASTKORPKRNLKADVEEFLALR 2377
QY 2216 --TPS-----QLSPGQSOQVQTTTQPIQPHSTSL-QIP-SOGOPQSQP----- 2256
Db 2378 KRTPSAGKAMDTPKPAVSDKKNINTFVETPVQKLDLLGNLPGSKRQTPKEKAELEDL 2437


```
QY 1157 VSIQDSSSEDMIVQNSNESISEQF-RTREODVEVLEPLKLCVLSGESTGNCEDRLPVKGT 1215
Db 1256 EKITRTASKNAADLEHPEPDLPSLRTRRNRVSY-----ATMGDHENSPV----- 1303
QY 1216 EANGKKPSQOKLEERPVNKKSDQIKLNTDKN-----NENRESEKK----- 1259
Db 1304 ---KEPVEQPRVTRKRLERELQEAATAVPTPRGRPPKTRRRADEBEENEAKAPETLK 1359
QY 1260 ---GORTSTFOINKDKPKIYLKGECLKEISRVVSGNVEPKVNNINKIIPENDIK-- 1314
Db 1360 PPEGWRSRPSOKTAAAGGP---QCK-----KGKNEPKVDATR---PEATTEVG 1401
QY 1315 -SLTVKESAIRPFINGVIMEDFNERNSETKSHLLSSDAEGNRYRDSLETLP----- 1366
Db 1402 POIGVKESMEP-----KAAEEAGSEQKR---DRKDA-GTDKNPPTAPVVEVVEKK 1449
QY 1367 ---STKESDSTQTTSPASCSPESNVNQVE-----DMEIETSEVKKVTSSPITSEE 1414
Db 1450 PAPEKNSKRGSRNSRLAVDKSASLKNVDAAVSPRCAAAQAGERESGVVAVSPEKSES 1509
QY 1415 -----BSNLSNDFIDENGLPIKNNENNGESKRKVTITEVTTMTSTVATESKTVIKVE 1467
Db 1510 PQKEDGLSSQLKSDPVDKPEKEDEVSAAGSPSEATOLAKQMELEQAV----- 1558
QY 1468 KGDKQTVVSSSTENCAKSTVTTTITVTKLSTPSTGGSDIISVKEQSKTVVTTVTDLSLT 1527
Db 1559 -----EHTAK-LAESASAAKADAP-----EGLAPEDRDKPAHOASETE-LA 1599
QY 1528 TTGGTLVTSMTVSKYESTRDKVKLMKFSRPK-----KTRSGTALPSYRKFTVKSTKKSIFV 1583
Db 1600 AAGSIINDISGEPE-----NFPAPPYPGESQTDLOPPAGAQAQLOPS----- 1642
QY 1584 LPNDLKLAKRGKIREVPYNNYNAKPAIDTWYPSPRPTEGITWYRLQTVKSLAGVSL 1643
Db 1643 ---EEGMDTDEAVSGILEATEASSRP-----PVNAPDPSAGPT-----DTKEAGNS- 1688
QY 1644 MLRLWLASLRWDDMAAKVPPGGGGSTRYETSETITTEIKRRDVGPGYIRFEYCIRKII 1703
Db 1689 -----SETSHSVPEAKGS-----KEVEITLVKKDGRQKTRSRKRKNWNKKVV 1732
QY 1704 CPI--GVPETPK---ETP-----TPQRKGLRSSALRPKRPTPKQTGP----- 1741
Db 1733 APVESHVPESNAQGESPAANEGETVQHPEAQEE-----KQSEKPHSTPQSCTS 1783
QY 1742 ---VLIETWABEELEWEIRAFARVEKE-----KAAQVEQAKKRLKQKPVIAFATST 1793
Db 1784 DLSKIPSTENSQSEISVEERTPTKASVPPDLPPPPQAPVDEEPQARFRVH--SIIESDP 1841
QY 1794 TSPTSSSTSTISPAQKVMVAPISGVTGTRKWLTTKVGSPATVTFQONKFNHQTATWV 1853
Db 1842 VTPSPDPSIPTLPSVTAALKSPPVASG-----GIP-----HOSPPTKV 1881
QY 1854 KQGSQNSGVQVQOKVLGIIP-----SSTGTSSQOTFTSFQPRATATVIRPNTSGSGGTT 1908
Db 1882 TEWITRQEPRAQSTPSPALPPDTKASDVDTSSSTLRKILMDPKYVSNATSVTSVITAI 1941
QY 1909 NSQVITGP---QIRPGMTVITPQOOSTL-----GKAIITPVWVQGE-----APQO 1952
Db 1942 AEPYSAAPCLHEAPPPVDSKKPLEEKTAPPVTNNSEIQASEVLVAADKEKVAAPVIAPK- 2000
QY 1953 VMTGIIIRQOPVSTAVSAPNTVS-STPGOKSLTSATST-SNIQSSASOPPRPQOQVKILTM 2010
Db 2001 -ITSVISMSPVIDLNSQKITLAKPAPQITGLVLSALTGLVNVSLVPVNAKLPVKGVS 2059
QY 2011 AQLRLTQGHGNGOGLTVVIOQG-GOTTGQLQLIPQGV-----TVLPFGPQOOLMQAAAMPN 2064
Db 2060 TTLKSLVSTPAGPVN---VLKGPVNVLTGPNVNLTPPNATVGTVNAAPGTVNNAASAVN 2116
QY 2065 GTVORELFTPLATTATTAATTTTTVSTTAAG-----TGEORQKSLSP-QMOVH 2111
Db 2117 ATASAVTVTAGAVTAASGGVTATGTGTVMAGAVIAPSTCKQKQASANENSFRHFGSMPVI 2176
QY 2112 QDK-----TLPQAQSSSV-----CPAK 2128
```

```
Db 2177 DDRPADAGSGAGLRVNTSEGVVLLSYSGQKTEGQRIASIKISQIPPSAMDIETFOQSVSK 2236
QY 2129 AQPOTAFSPARPQOTQPSPA-QPEVQTOPE--VQTQTTVSSHVSPSAOPTHQSSKQP 2185
Db 2237 SOVKPDSVTASQPSKGPQAPAGYANVATHSTLVLTACTYNASPVISSVKADRSLEKPE 2296
QY 2186 VAAQSQSQSNVQGS-----PVRVQSPSQTRIRPS---TPSOLSPGQSQSVQVTT 2231
Db 2297 PIHLSTVSTPVTQGGTGVKVLTOGINTPPVLVH--NQLVLTSPSIVTTNKKLADPVLTKIETK 2354
QY 2232 TSQPI----PIQHTSLQIPSQGQPOQPOQVQSQTTLSSGQTL-NQVSVSSPSRPQLOI 2286
Db 2355 VLQPANLGSTLTPHPHPALPS-----KLPTENVHVPSPGSPADRTVSHLAAAKLDA 2406
QY 2287 QOPOPO---VIAVPLQOQOVQLSQIQSVVQAQQAQSGVPPQIKLQPLQIQOQSSAVQ 2343
Db 2407 HSPRSPGPGSPSPFRASHPSSTASTALSTNATVLAAGIPVPOFIS---SIHPGQSVMP 2463
QY 2344 THQTONVVYVQAASVQEOLOQ-----RVQOLRDQOQKKKQOQIETKREHTLQAS 2391
Db 2464 PHSITQTVSLSHLSQGEVMMNTPTLPSITYSIRPEALHSRAPLQPOQIEVRAPQRASTP 2523
QY 2392 NQSEIIQKVVMKH---NAVIEHLKQKSMTPAEREENORMIVCNQVMKYILDKIDKEEK 2448
Db 2524 QPAPAGVPALASQHPPEEVEHYHLPVAKATAPVQSE-----VLVMQSEYRLHPYTVPRDVR 2579
QY 2449 QAAKKRKREESVEQKRSKQNAKLSALLFKHKEQLRAEILKRLALLOKDLQIEVQOELKR 2508
Db 2580 IMVHPHTAVS-EQPRADGVVYVP-ASKAPQO----- 2611
QY 2509 DLTKTKKEDLMQLAQAATAVAAAPCPVTPVLPAAPAPPPPPPGVQHTGLLSTPT 2564
Db 2612 -----PGKEAKTDPKAAAPTTPAPVPV-----VPVLPAPAPAPHGEARILTVTPS 2658

RESULT 8
US-09-919-172-98
; Sequence 98, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919.172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 98
; LENGTH: 3256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 2700132CD1
US-09-919-172-98

Query Match 2.88; Score 403; DB 10; Length 3256;
Best Local Similarity 18.5%; Pred. No. 9e-12;
Matches 523; Conservative 371; Mismatches 1011; Indels 916; Gaps 130;

QY 339 IIEEDTENENEKKTWYSTKVQLAELIDCLDK----DYWEAELCKILEEMREEHIMDI 394
Db 432 VLPTEIHNPEFTLWLTQVERKIQKDSLSPKELGTAGQMSCGLPGLSS-----VDI 486
QY 395 TE--DLTNKARG-----SNKSFLAAANEETILE-----SIRAKGDIIDN-----VKSPPE 436
Db 487 NNFGDSINESEGIPLKRRRSFSGHLRPELFDENLPNTPLKRGCAPTKRKSLVMMHTPPV 546
QY 437 TEKDKNETENDSKDAENKRE---EFEDQSL-----EKSDDDKTP-DDDPEQKSEVGFDFS 488
```

[illegible]


```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-649A-425

```

Query Match 2.9%; Score 418; DB 15; Length 4019;
Best Local Similarity 17.4%; Pred. No. 2e-12;
Matches 492; Conservative 320; Mismatches 995; Indels 1018

Qy	254	CRVCHKL--GDILLCETCSAVHLECVKPPLEEVPEDEMOECVCAHVKPGVTCVAB1	310
Db	65	CEACGRATDPGRLLLLCCDDISVHTYCLDPPJOTVPKGWKCKWCWCRHCATSNAGLR	124
Qy	311	QKNKPYIRHBPICYDRSRKRYPLNRRLLITEEDTENENKKIWIYSTKVOLAELIDCLDK	370
Db	125	EWQNNYTQACPA-SLSSCPVCYRNY--EEDL-----ILQCROC	161
Qy	371	DYWEAEALCKTLEEMREEIHRHMDITEDLNKAGSNKSFLLAAN-----EELLESIR	422
Db	162	DRWHAVCQNL-NTEEEVENADIGFDCS-----MCRPYMPASNVPSDCCESSLVAIV	215
Qy	423	AKGGDIDNVKSPB-----TEKDKNETENDSKDAENRREEFEDQSLSEKSDDKT-----	471
Db	216	TKVKLDPPTTYTQDGVCLTFESGMTOLOSLUTVTVPRKRKSKPKLKLKIINONSAVILQTP	275
Qy	472	PD-----DDPQGRKSEVGDPKSEKSGELSES PGAGKAGSGASTRIITRLRNPDS	520
Db	276	PDQTSEHSRDMDDDSREGELMDCDGKSE-SSPEREAVDDTEGVGEGTDGKVRKRKP--	332
Qy	521	KLSQLKSQQVAAAHAANKLFKGEKVELVYVNSOGETISRLSTKKEVINK---GNINNYFKL	577
Db	333	-----YRPGIGFMVQRORSRTQGQTKRSVIRKODSSGSGISBOLPC	372
Qy	578	QBGCKYRVYHNQYSTNSFALNKHQHRHEDHKRRHLAHKFCLTLPAGEFKWNGSVHGSKVLT	637
Db	373	RDDG-----	376
Qy	638	ISTLRLTITOLENNIPSFLLHPNASHRANWIKAVOMCKSPREFALALALECAVKPVVM	697
Db	377	-----	376
Qy	698	LPIWREFLGHTRLHRMTSIRREEKEVKKKEKQBE--EET---MQOATWVKYTFPVKH	751
Db	377	--WSEQLPDLVDSESV-TESTEKIKYRKRKNKLEETFPAYLQEAFFGKDLDTSR	432
Qy	752	QVMKQGEEXRVTGYCGWSWIKTHVYRVPFLPGNTNVNRYKSLGTEKKNMDDESD	811
Db	433	Q-----SKISLDNISEDGNQLLYKTNN-----TGFLDPSLD--P	465
Qy	812	KRGRSRSPKKIKIIPDSEKDEVKSDAAKAGADQNMEDISKITEKKODVVELLSDSDKP	871
Db	466	LLSSSAFTK-----SGTHGPADDPDLADISEVL-NTDDOILGIISDDLAKS	510
Qy	872	CKEEPM-EVDDDMKTESHVNCQESSQVDVNVNVSFGFHLRTSYKKTKSSKLDGLLERIK	930
Db	511	VDHSDIGPVTDDPSLPQPNYQSS-----RPLSEEQJLDGLL-----	547
Qy	931	QFTLEEKORLEKIKLEGGIK-IGKTSTNSSKNLSESPVITKAKEGQCSMSRQEQSP--	987
Db	548	-----SPELDKWDTGAILGLKYIKPELGGKQVEDLFTAVLSPANTQPTLPQPPPTQ	601
Qy	988	-----NANNDQPEDLIQCSQSDSVLRMSDP-----SHTTNKLYPKDRVLDDVSIR	1034
Db	602	LLPIHNQDAFBRWPLMNLGSSPHLPHNSLPPGSGLGTFSAIQAQSSVPOAR--DKNSAF	659
Qy	1035	SPETKCPQN-----STENDIEKVSJLASRGQEPKSKTKGNDFFIDDSKLASADDIG	1088
Db	660	NPMASDPNNSWTSAPTVEGE-----NDTMSNAQRSTLKWKE-----EALGEMATVA	707
Qy	1089	TLICKN-KKPLQIESDITVSSSKSALHSSVPKSTNDRDATPLSRAMDFEGKGLCD--SE	1145
Db	708	PVLYTNIINPNLKEEFPDWTTRVKQ-IAKLHRKASSQERAPYVQKARDNRAALRINKVQM	766
Qy	1146	SNSTLSENSDVTSIQDSSE-----EDMIVONSNESISEQFTREQ-----DV	1187

Qy	1996	-SQPPRPPQGVK-----LTMQIQTQTHGHNQGLTVVI	2030
		:	
Db	1745	DDDPVKELVDKLEGVEFKLDDLENLNLDTEGKVVLEDTLDNLETDPNLDLAL	1804
Qy	2031	QSGOQTGTQQLLP-----QGVTVLPGPQQLMQAA	2061
		:	
Db	1805	R-----SGEFDIIATYDPELDMGDKKSMNEELDLPIDDKLNQCVSVEPKKQE	1854
Qy	2062	MPNGTVQREFTPLATTATTATTTTSTVTAAGTGEORQSKL-----SPQM	2108
		:	
Db	1855	-----ENKILVLSDKHSPQKKSIVTEVTKTEVLSPNKSVESKETEKNDEKNDVTPCS	1909
Qy	2109	QVHQDKTLPAPQSSSVGPA-----KAQPQTAPQARPOPQTPQSPAPQEVQTOPEVQ	2161
		:	
Db	1910	QASAHSDLNDGKETSHPCDPDLFEKTRNRETAGPSAN-VIQAOSTQLPAQDVINSCGITG	1968
Qy	2162	TQTTVSHVPSAQPT---HAQSSKPOVAAQSQPOSNVOGQSPVRVQSPQTRIRPSTPS	2218
		:	
Db	1969	STPVLSSLANEKSDNSDIRPSGPPPTLPAPSNHVSLSLPPF-----IAP----	2015
Qy	2219	QLSPGQ-----SQOVQTTTSPQIPIQHTSLQIPSQ-----GQPOSQ--	2255
		:	
Db	2016	---PGRVLDMAMNSNVTVSVRNHVFSQGVQVNPGL--IPGSTVNHSLGTCKPATQTG	2069
Qy	2256	PQV-QSSTOTLSGSQTL---NQVSVSPSRPQIQOQPQPVTAVPOLQOVQVLSQIQS	2311
		:	
Db	2070	PQTSQCTSSMSGPQQLMIPQTLAQONRRERPLLEEQPLLDLQEROEQOQQQRMQA	2129
Qy	2312	QV-----VAQIQAQOS-GVPQOIKLQLPQIQOSS	2340
		:	
Db	2130	MIRORSEFPFNIDFADITPIMKAKMVALKINKVMAQNMLGMPVMVSRRFFPMQGVVT	2189
Qy	2341	AVQTHQTVN---VTVQAASVQEOQLRVQO-----LRDQOKKKQOQIQIEKREHTLQA	2390
		:	
Db	2190	GTQNSEQNLGPQAIPODGSITHOISRPNPNFGPGFVNDQSQRKQYEEWLQ-----	2240
Qy	2391	SNQSEIITQKVYMKHNVAIEHLKOKKSMTPAEREENORMIVCNOVMKYILDKTDKEEKQA	2450
		:	
Db	2241	-ETQOLLQMQOKYLEEQIGAHRRKSKKALSQAKR-----T	2273
Qy	2451	AKKREESVEOKRSKQNAIKLSALLFKHEQLRAEILKKRALLDKDLQIEVQOEEELKRD	2510
		:	
Db	2274	AKKAGREFPEDEAQLKHVTVEQSNQVQLEQIRK-----QOKEHALLIEDY	2320
Qy	2511	KIKKEDLMQLAQATAVAAP-CPPVTVPVLPAAPPSPPPPPGV-----QHTGLLST	2562
		:	
Db	2321	RIKQOQ---QCAMAPPTMPSVQOPPLIPGATPTTMSQPTTFMVPVPOOLQHOQHTTVISG	2377
Qy	2563	PTLPV	2567
Db	2378	HTSPV	2382

```

RESULT 6
US-10-144-649A-425
; Sequence 425, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITILE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; TITLE OF INVENTION:
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
;

```

; ORGANISM: Homo sapiens
US-09-854-133-425

Query Match 2.9%; Score 418; DB 10; Length 4019;
Best Local Similarity 17.4%; Pred. No. 2e-12;
Matches 492; Conservative 320; Mismatches 995; Indels 1018; Gaps 103;

254	QY	CRVCHKL---GDLLCCTCSAVTHLCEVCBPPLBEEVPEDEWQCEVCVAHKVPGVTDCAVEI	310
65	Db	CEACGRATDPGRLLLLDCCDISYHTYCLDPLQTPVPGWKWKVCWCRHCGATISAGLRC	124
311	QY	QKNKPYIRHEPIGYDVSRRRYWFLNRRLLITEEDTENENEKKIWIYSTYKVLQALAEIDLCLDK	370
125	Db	EQWNNYTQCAPCA-SLSSCPVCYRNYR---EEDL-----ILQCRQC	161
371	QY	DYWEALCKTILEMREBEIHRMDITDITNKARGSNKSFIAAAN-----EBILSIR	422
162	Db	DRMHAVCQNL-NTEEBEVENADIGFCS-----MCRPYMPASNPVSPDCCESLSVAQIV	215
423	QY	AKKGIDINVKSPPE-----TEKDKNETENDSKDAEKNREEFEDQSLEKSDSDDKT-----	471
216	Db	TKVKELDPPTYTQDGVCLTSEGMTQLOSLTVVPRKRKPKLKLIIQNONSVAVLQTP	275
472	QY	PD-----DQPEQKSEVDFKSEKNGELSESPGAGKASGSTRITILRNPDS	520
276	Db	PDIOSEHRDGMDDSGELMLDCDGKSE--SSPEREAVDDETKGVEGTGDKVKRKRPP--	332
521	QY	KLSQLKSOQVAAAHEANKLFPKEGKEVLVYNSQGEISRLSTKKREVIK-----GNINNYFKL	577
333	Db	-----YRGIGGFMYQRKSRTOGQGTKKRSVIRKDSGSGISEQLPC	372
578	QY	GOBCKYRVYHNOYSTNSFALNKHQHRDHDKRRHLAHKFCLLTPAGEFKWNGSVHSGKVL	637
373	Db	RDG-----	376
638	QY	ISTLRLTITOLENNIPSSFLLHPNWSHRANWIKAVQMCSPREPALALALECAKVPVM	697
377	Db	-----	376
698	QY	LPIWREFLHTRLHRMTSIEREEKEVKKKEKQEE--EET---MQOATWVKYTFPVKH	751
377	Db	---WSEQLPDLVDESYSV-TESTEKIKKRYRKRKNKLEETFFPAYLQEAFFGKDLLTSR	432
752	QY	QVWKQKGEERYTGYGGWSHISKTHYRVFKPLPGNTNVNYSKLEGTKNMNDENDES	811
433	Db	Q-----SKISLDNSEDGAQLLYKTNMN-----TGFLDPSLD--P	465
812	QY	KRKCRRSPKKIKIEPDESEKDEVGSDAAKAGADONEMDISKITEKKQDVVELLSDSDDKP	871
466	Db	LLSSSSAPTK-----SGTHGPADPLADISEVL-NTDDDLIGIISDDLAKS	510
872	QY	CKEPM-EVDDDMKTBHVNCQSSQDVVNVNVEGPHLRTSYKKTKSKKLDGLLERRIK	930
511	Db	VDRSDTGPVTDPSLPPQPNVNSS-----RPLSEEOQLDGIL-----	547
931	QY	QFTLEEKORLEKIKLEGGIKG-IGKTSTNSKKNLSPEVITKAKEGQCSMSRQEQSP--	987
548	Db	-----SPELDKAVTQGAIIKGLKYKIPELCGKDVEDLFTAVLSPANTQPTLPQPPPTQ	601
988	QY	-----NANNQDQEDLIGQCSQSDSSVLRMSDP-----SHTNKLKPKDRVLDDYSIR	1034
602	Db	LLPIHNOQAFSRPLMNLGLSSPHLPHNLSLPGSGGLGTFSATAOSSVYPDAR--DKNSAF	659
1035	QY	SPETKCPQN-----SIENDIEEKVSDLASRQGEPTKSYKTGNDFDIDSKLASADDIG	1088
660	Db	NPMASDPNNSWTSAPTEGE-----NDTMSNAQRSTLKWEK-----EALGEMATVA	707
1089	QY	TLICKN-KKPLIOESDTIVSSKSLAHSSVPKSTNDORDATPLSRMDFEGKLGCD--SE	1145
708	Db	PVLYTNINFNPLEEPDWTTRVKQ-IAKLWRKASSOERAPYVQKARDNRAALRINKVQM	766
1146	QY	NSFTLENSDTSVDSQDSE-----EDMIVQNSNESISEQFRTREQ-----DV	1187

Db	767	SND5MKKQQQODSDTDPSSRIDSELFKDOPLKQRESEH--EQEWKFKQQQMRQSKQQAQKTEAT	825
Qy	1188	EVLEPKECE-----LVSGESTGNCEDRLPVKGTKEANG-KPSQQ-----	1225
Db	826	QKLEQVNEQEQOQQOQFGSQHLLLVQSGSDTPSSGIGSPLTPQPGNGNMPAQSFHKELF	885
Qy	1226	-KKLEBRPVNKCSDQIKLKNT-----DKNNENRESEKKQORTSTFQINGK	1271
Db	886	TKQPSTPTSTSSDDVFVKQAPPPPPAPRPIQDLSQAQTSQPPSPQVPSFGSSNSR	945
Qy	1272	DNKG-KIYLK-----CECLKEISESVGNSVEPKVNNINKI	1307
Db	946	PPSMDPIYAKWGTPRPPVGHGSSRRNSAAPVENCPTLSSVSRPLQMN-----ETTANRP	1001
Qy	1308	IPENDI--KSLTVKESAIRPFINGDIVMEDFERNESSETKSHLLSSSDAEG-----	1356
Db	1002	SPVRDLCSSTTNDPYAKPDPDTPRVVMTDQFKSLGSLRSPVYSEQTAKGPIAAGTSDH	1061
Qy	1357	-----NYRDSLETLPSTKESDSTQTTTPSASCPS-----NSVNOV-E	1393
Db	1062	FTKPSRADVFQORIPDSYARPLLT-PAPLDSGGPFKTPMQPPPSQSDPYGSVSOASR	1120
Qy	1394	DMETETSEVKVTSPSTSEESNLSNDFIDENGLPINKNENVENGSKRTVITEVTMT	1453
Db	1121	RLSVDPYERPALTPRPDINFESH-NQSNDDPSQP--PLTPHPAVN-----ESFAHP	1167
Qy	1454	STVATESKTVIKVEGDKQTVVST-----ENCAKSTVTTTTTTVTVKLSTPSTGSGVDII	1508
Db	1168	SRAFSQPGTISRPTSQDPYQPPCTPRPVDSYSQSGTARSNTDPSYQPGTPRPTVD	1227
Qy	1509	SVKEQSKTVVTTVTDLSLTTGGTGLVTSMTVSKYESTRDKVKLKMFRPKKTRSGTALPS	1568
Db	1228	PYSQOPQTPRPSTQDLEFVTP---VFNQRHSDPYAHPGTPRPGISVPYSPQPPATPRR	1283
Qy	1569	YRKFTVSTKKSIFVLNDDLKLRGKIGREVPYFN--YNAPALDIWYPSRPTFGI	1626
Db	1284	ISEGTFKSSMTRPVLMNQD-----PFLQAQNRGPAL---PGLVLRP-----	1323
Qy	1627	TWRYLRQTVKSLAGVSLMLRLMLASLRWDDMAAKVPPGGGSTRPTETSETITTEIIKRR	1686
Db	1324	-----	1323
Qy	1687	DVGPGYGRFEYCIRKKICTPGVPTKETPTPQKRLGRS--SALRPKRPETPKQTPGVII	1744
Db	1324	-----PDTCSQTPRPPGPGLSDFSRVPSAARDPYDOSMTP	1361
Qy	1745	ETWVAEELELWEIIRAFAERVEKEKAQVEOAKKRLQOKP---TVIARTSTSPSTST	1801
Db	1362	RS-----QDSFGTSGTAHDVADQPRPGSEGFECASSNSPMUISQG	1401
Qy	1802	STISPAQKVMVAPISGSVTTGTCKMVLTTKVGSPATVTFQONKNPHOTFATWVGQGSNSG	1861
Db	1402	QOFSGVSQ-LPGPVPTSGVTD-----QNTVNQAQADTEKLQRQOKLRE	1444
Qy	1862	VV---QVQKVGLGIIPSTGTS-----QOFT-----	1885
Db	1445	IILQOQOQKKIAGRQEKGSODSPAVPHPGPLQHWQPNVNAQFTRPPPPYGNIRSPVAP	1504
Qy	1886	-----SFQPRATVTRIPN-----TSGSGGTTSNS-----QVITG--	1915
Db	1505	PLGPRYAVFPKQDRGYPDPDVASMGMRPHGRFPGFGSHGTMPSQBRFLVPPQOIOGSG	1564
Qy	1916	--POIRPGMTV-----IRPPL-----QOSTLGRKAI-----	1939
Db	1565	VSPQLRSVSDMVRPLNNSQMNPNVGLPQHFSQSLPVQOHNILGOAYIELRHAPDGR	1624
Qy	1940	-RTPVMVQPGA-----	1957
Db	1625	QRLFSAPPGGSVVEASNLRHGNFIRPDPGPRHTDMRRPPQGLPNQLPVHPDLEQVP	1684
Qy	1958	-----IRQPVSTAVSAPNTV-----SSTPGOKSLTSATSTSNIOSSA-----	1995
Db	1685	PSQOEQGHSHSSSWMARTLNHPILGGBFSAPLTSVPSETSTSNLQITTTQPSDGBLEKL	1744

875 MPCVVLTRYKEGKVIDHTPV-----EKLAKALDNDTVKSSALDQKLOVSQT 922
Qy
165 TFGPADLK--DSVNSTLYFDGM-TWPEVLRVYCESD---KEYHHVLPYQEAEDYDGPV 218
Db
923 EPAKSDLSKLESVMKVPKGLSHVEVY---EKEGRLKARKHLKBPQPAD---GVS 974
Db
219 ENIKIVLOFLVDGFLNTNARELMSE---GVIOYDHGRVCHKLDGDLCCETCSAVYHL 275
Qy
975 AVDLKLEARKRFADSNLKAQKQPEVKKSSPEMED-ARVLSSKOPDVSRRVILLREG 1033
Db
276 ECVKPPD--EEVPEDSQCEVCVAHVPGVTDGVAIOKNKPYIRHEPICYDRSRKYWF 333
Qy
1034 EAERKPVREILAKRESKIKLDRLNTVAPKQC-QELAS-----ISVGSGRSPSD 1083
Db
334 LNRRLL--IIBEDTENENEKIWIYTKVQLAEILIDCLDQDYWEAECKILEEM---REEL 388
Qy
1084 LOARLGLAGESVENOEVOSKPIPSKPOLKQL-----QVLDDQGPREDV 1129
Db
389 HR-HMDITDLTKARGSNKSFLLAAANEETILESIRAKG-DIDNVKS---PEETEKDNE 443
Db
1130 RKNYCSLRDTPPERKSGOESHVNTTEKI-----GIDIDHTOSYRKQMQSRKQ 1181
Db
444 TENDSKDAE-----KNREFEQSL-----EKSDDKTPDDPEQOK-SEVGDF-----486
Qy
1182 MEMEIAKSEKFGPKDQDVEYERRSLVHEVGKPPQDVTDSPPSKKRMDHVDICTKR 1241
Db
487 -KSEKNGELSE-----SPGAKGA-----SGSTRITILRNP--DSKLSQLKSQ 529
Qy
1242 ERNYSRQIASEDSERTGGSVRHGSFHEDEPIGSPRLSVKGSQVDEKVLPSYNT 1301
Db
530 VAAAHAANKLFEGEVL-----VYNSOGEISRLST-----K 562
Qy
1302 VRESLKFNYDSSRRQOMADMAKIKLSVLNSEDELNRWDSQMDAGREDVDFPNSIIK 1361
Db
563 KEVIMKGNINY-----FKLGQGYRVYHNOYSTNSFALNKH-----QHRED-----605
Qy
1362 RDSLRKRSVRDLEPGEVPSDSEGE-----HKSHSPRASALYESSRSLSFLLRDRDKLRE 1417
Db
606 HDXR-----RHLAHRECL-----TPAGEFKWGSVHGSKVLTISTLRLTLTQLENNIPSS 655
Qy
1418 RDERLSSSLERNKFYSFALDKTTP-----DTKALLERAKSLSSREEN---WS 1463
Db
656 FLHPNWAHSHRNKIKAVQMSKIPREFALALALECAVKPYVMLPIWREFLGHTRHMTS 715
Qy
1464 FL--DWDSRFANFN-----NKDE-----KVDSAPRPI---PSW-YMKKKKIR---1501
Db
716 IEREKEKVKKKKQBEETMQOATWVKTFPVKHQVWQKGEERYVIGYGSWSLTKT 775
Qy
1502 TDSEKMDKKEHKEEQERQE-----L 1525
Db
776 HVYRFVPLPONTNVYRKSLGKTKNMNDENDESRRKCSRSPKKIKIEPDSE--KDEV 833
Qy
1526 PASRF-----LHSEIFEQDSKRLQHLERK---EEDSDFISGRI 1560
Db
834 KGSDAARKGADQNE-----MDISKITEK---KDQVDKELLSDS-----DKPC 872
Qy
1561 YGKQTSSEANSTTDSIQEPVVLPHSRFEMELTRMOQKEKDKQPEVEKQEDTENHPKTP 1620
Db
873 KEPEMEVDDDMKTESHVNCQESSQVDVNVVSEGHRTSYKKTKSKSLDGLL---ERRI 929
Qy
1621 ESAPENKDSLEKTPPSVG---PPSVTVVLTESA---PSALEKTTGDKTVEAPLVEEKT 1674
Db
930 KOFTLEEK-----ORLEKIKLEGIGIKGTSTNSSKNLSESPVITKAKGQ 977
Qy
1675 EPATVSEAKPASPAPAPVEQLQVDPG-----ADPKEAAMP--AGVEBGS 1724
Db
978 -SDSM-----ROBQSPNANDQPEDLIQOCSQSDSVLRMSDFSHNTNKLKYPK----DRVLD 1029
Qy
1725 GDQPPYLDKAPPTPCAFSQAESNVDPDPDSTQPLSKPAQKSEANE--PRAEKPDATAD 1782
Db
1030 ---DVSIR---SPETKCPKQNSIENDIEEKVSDLASRGQBTYSK-----1068
Qy
1783 AEPDANKAEAAEPSPASEDLEVD-----PPVAAKDKKPNKSKRSKTPVQAAAVSIVE 1837
Db

1069 ---TKGNDEFFIDDSKLASADDIGTLICKNNKPLIOESDPTIVSSSKSALHSSVPKSTNDR 1125
Qy
1838 KPVTRKSE-RIDREKLRNSNPRGEAQKLELLEKMAEKITRTASKNSAADLHPE-----1891
Db
1126 DATPLSR-----AMDFEGKLGCDSSNSTLENSDTPV-----SIQDSSEEDMIVON 1171
Qy
1892 PSILPSTRRRNVRSVYATMG-DHENRSPVKEPVEQPRVTRKRLERELQEAATAVPTPRR 1950
Db
1172 SNESISEQRTREQDVEVLEPLKCELVSGESTGNCEDRLPVKGTEANGKKPQQOKLE-E 1230
Qy
1951 GRPPKTRRADDEEENEKEPAE-TLKPPPEGRSPRSQKTAAGGGQGGKKNPEKVDAT 2009
Db
1231 RP--VNKSDOIKLKNNT-----DKNNENRESEKKGORTSTFOI-----1268
Qy
2010 RPEATTEVGPGVIGKSSMEPKAAEEASEQKDRKDACTDKNPETAPVEVEKKPAP 2069
Db
1269 -----NGKDNKPKIYL-KGECLKEISES---RVVSGNVPEPKVNNINKIIPENDIKSUT 1317
Qy
2070 EKNSKSKGRSRNSRLAVDKSASLKNVDAASPRGAAQAAGERESGVAVSPEK---SESP 2127
Db
1318 VRESATRPFGINDVI-----MEDFNERNSSSETKSHL-----LSSDAE 1355
Qy
2128 QKEDGLSSQLKSDPVPDPKEPEKEDVSAGSPSEATOLAKQMELEQAVHIAKLAEASAS 2187
Db
1356 GNYR-DSLETL-PSTKESDSTQTPPSASCPESNSVNOV-----TESKTVIAXEKGDKQTVVSSTE 1479
Qy
2188 AAYKADAPGLAPEDRDKPAHQASETELAAGSIINDISGEPENFPAPPYPGESOTDL 2247
Db
1393 -----EDME-----TETSEVKVTSPTITSEESNLSNDFIDENGLPIN 1431
Qy
2248 QPPAGAALQPSSEGMETDEAVSGILEATAATESRPPVNPAPDPSAGPTDTKEARNSSE 2307
Db
1432 KNEVNGESKRKTVITVETMTSTVA-----TETTVIAXEKGDKQTVVSSTE 1479
Qy
2308 TSHSVPEAKGSKEV---EVLVRKDKGRQKTRTSRRNRNTNKKVAVPESHVSPESNOAQGE 2365
Db
1480 NCAKSTVTTT-----TTVTYKLTSTPGSGVDIISVKDESKT 1516
Qy
2366 SPAANEGTIVQHPPEAQEEKQSEKPHSTPPQSCSTSLSKI--PSTENSSQEISEVERTPT 2423
Db
1517 VVTTTVTDSLTTTGGTLVTSMTVSKYSPDKVKLMKFRPKKTRSGTALPSYRKVFVTKS 1576
Qy
2424 --KASVPDPLPPP---POPAPVDEEPOARFVHSIIESDPVTPPSDPSIP-----2468
Db
1577 TKKSIFVLPNDLKLKLA---RKGGIREVPYFNNAKPALDIMPYPSPRPTFGITWYRLQ 1633
Qy
2469 -----IPTLPSVTAAKLSPPVASGII-----PHQSP-PTKVTEWITRQE 2506
Db
1634 TVKSLAGVSLMLRLLWASLRWDDMAAKVPPGGGSTR-----TETSETEITTTTETIKRRD 1687
Qy
2507 EPRAQSTPS-----PALPDDTKASDVDTSSSTLRKILMDPKYVSATSSTSV-----2554
Db
1688 VGPYGRFECYIRKILCIPGVPETPTPQKGLRSSALRPKRPETPKQTPGVILETW 1747
Qy
2555 -----TTAIAEPVSAAPCLHEAPPP-----PVDSSKKPLEETAPV-----2590
Db
1748 VAEELLEWEIRAFARVEKEKAQAV-----EQOAKRLEQOQKPTVI 1789
Qy
2591 TNNSEIQASEVLVAA---DREKAVPVIAPKITSVISRMPVSIIDLENSOKITLAKPAPQTL 2647
Db
1790 ATSTTSPTSTTSTISPAQKVMVAPISGVSVTGTVKVVLTGKVGSPATVITQQKNFHQTF 1849
Qy
2648 TGLVSALTGLVNVSLVPV-NALKGPKVGSVTT-LKSLVSTPAG-PVNVL-----2693
Db
1850 ATWVKQGOQS--SGVQV---QOKVLGIIPSSGTSSQOFTTFQPTATVTRPN--TS 1901
Qy
2694 -----KGPVNLVTPGVNLTTPVNAVTVGTVNAAPGTVNAASAVNATASVTTAGAVTA 2748
Db
1902 GSGGTTNSQVIT-----GPQIRPGN-----1922
Qy
2749 ASGGVATTGTVTMAGAVIAPSTCKQORASANENSREHPGSMPIVDDRPADAGSAGLRV 2808
Db

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (286)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-816

Query Match 10.3%; Score 1474; DB 9; Length 328;
Best Local Similarity 94.8%; Pred. No. 4.6e-68;
Matches 289; Conservative 0; Mismatches 14; Indels 2; Gaps 1;
QY 406 NKSFLAAANEILLESIRAKKGDIDNVKSP--EETEDKNETENDSKDAEKNREFEFQSL 463
DB 12 NVSILSGGTNEILLESIRAKKGDIDNVKSPGTGEETEDKNETENDSKDAEKNREFEFQSL 71
QY 464 EKSDDKTPDDPEQKSEVGDGPKSEKNGELSESPGAGKAGSGSTRITPLRNPD SKLS 523
DB 72 EKSDDKTPDDPEQKSEVGDGPKSEKNGELSESPGAGKAGSGSTRITPLRNPD SKLS 131
QY 524 QLSKQVAAAHAANKLFKEGKVLVNVNDSOGEISRLSTKKEVIMKGNINNYFKLGQEGKY 583
DB 132 QLSKQVAAAHAANKLFKEGKVLVNVNDSOGEISRLSTKKEVIMKGNINNYFKLGQEGKY 191
QY 584 RVYHNOYSTNSFALNKHQHREDHDKRHLAKHFKCLTPAGEFKWNGSVHGSKVLTIPTLRL 643
DB 192 RVYHNOYSTNSFALNKHQHREDHDKRHLAKHFKCLTPAGEFKWNGSVHGSKVLTIPTLRL 251
QY 644 TITOLENNIPSSFLHPNASHRANWIKAVQMSKPREFALALILECAVKPVVMLPIWRE 703
DB 252 TITOLENNIPSSFLHPNASHRANWIKAVQMSKPREFALALILECAVKPVVMLPIWRE 311
QY 704 FLGHT 708
DB 312 SLGHT 316

RESULT 2
US-09-729-835-80
Sequence 80, Application US/09729835
Patent No. US20010016647A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 29 Human Secreted Proteins
FILE REFERENCE: P2015p1
CURRENT APPLICATION NUMBER: US/09/729,835
CURRENT FILING DATE: 2000-12-06
PRIOR APPLICATION NUMBER: 09/257,179
PRIOR FILING DATE: 1999-02-25
PRIOR APPLICATION NUMBER: 60/056,270
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: 60/056,271
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: 60/056,247
PRIOR FILING DATE: 1997-08-29
PRIOR APPLICATION NUMBER: 60/056,073
PRIOR FILING DATE: 1997-08-29
NUMBER OF SEQ ID NOS: 128
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 80
LENGTH: 238
TYPE: PRT
ORGANISM: Homo sapiens
US-09-729-835-80

Query Match 6.9%; Score 982.5; DB 9; Length 238;
Best Local Similarity 78.8%; Pred. No. 5.1e-43;
Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;
QY 2563 PTLPLVASQKKRREEK-----DSSSKSKKKMISTTSKTKK-----DT 2601
DB 3 PTPP-----KTPYDESFYIGCDLCTNWHGECVGITEKEAKKMDVVICNDCKRAQEGSSE 58
QY 2602 KLYCICKTPYDESKFYIGCDRCQNWYHGRVCVGLQSEALIDEYVCPQOSTEDATWLT 2661

DB 59 ELYCICRTPYDESOFFYIGCDRCQNWYHGRVCVGLQSEALIDEYVCPQOSTEDATWLT 118
QY 2662 PLTEKDYEGKRLVLSLQAHKMAWPFLEPDPNDAPDYGVVKEPMDLATMEERVQRRYY 2721
DB 119 PLTEKDYEGKRLVLSLQAHKMAWPFLEPDPNDAPDYGVVKEPMDLATMEERVQRRYY 178
QY 2722 EKLTFFVADMTKTFDNCRYNPNSDSPFYQCAEVLESFVQKLGKFKASRSHNNKLOSTAS 2781
DB 179 EKLTFFVADMTKTFDNCRYNPNSDSPFYQCAEVLESFVQKLGKFKASRSHNNKLOSTAS 238

RESULT 3
US-10-177-293-423
Sequence 423, Application US/10177293
Publication No. US20030124128A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 423
LENGTH: 3664
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-423

Query Match 3.0%; Score 430.5; DB 15; Length 3664;
Best Local Similarity 18.5%; Pred. No. 4.1e-13;
Matches 550; Conservative 414; Mismatches 1028; Indels 989; Gaps 140;
QY 7 EEDGDAAEETQDSEDDE-----EDMEEDDDSDYPEEMEDDDDDASYCTESSFRSHS 59
DB 787 KNEKTKERTFDERVERERRRLIRKEKVKDKTDKQ-----822
QY 60 TYSSTPGRRKPRVHRPRSPILE-----EKDIPLEFPKS-----SEDLMVPNEHI 104
DB 823 -----KRGKGVHSPSSQSSSETDQENEREQSP-EKPRSCNKLRSREKADKEGIAKNRLEL 874
QY 105 MNVIAIEVLNRGCTVLRSLPFRFEFCALVSOEQCTLMAEMHVLLKAVLREDFDNT 164

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 01:17:18 : Search time 105.119 Seconds
(without alignments)
3939.095 Million cell updates/sec

Title: US-09-698-295-10
Perfect score: 14333
Sequence: 1 MYSEEEEDGDAEETQDSE.....KLKGFKASRSHNNKIQSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 14893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap:*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap:*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap:*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap:*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap:*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap:*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:*
9:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap:*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pap:*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pap:*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap:*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap:*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap:*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pap:*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap:*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap:*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap:*

ALIGNMENTS			
RESULT 1			
US-09-925-297-816			
; Sequence 816, Application US/09925297			
; Patent No. US20020081659A1			
; GENERAL INFORMATION:			
; APPLICANT: Rosen et al.			
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies			
; FILE REFERENCE: PAL05			
; CURRENT APPLICATION NUMBER: US/09/925,297			
; PRIOR FILING DATE: 2001-08-10			
; PRIOR APPLICATION NUMBER: PCT/US00/05989			
; PRIOR FILING DATE: 2000-03-08			
; PRIOR APPLICATION NUMBER: 60/124,270			
; PRIOR FILING DATE: 1999-03-12			
; NUMBER OF SEQ ID NOS: 928			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 816			
; LENGTH: 328			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: SITE			
; LOCATION: (170)			
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids			
; NAME/KEY: SITE			
; LOCATION: (172)			
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids			
; NAME/KEY: SITE			
; LOCATION: (174)			
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids			
; NAME/KEY: SITE			
; LOCATION: (178)			
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids			
; NAME/KEY: SITE			
; LOCATION: (183)			
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids			
; NAME/KEY: SITE			
; LOCATION: (269)			

SUMMARIES			
Result No.	Score	Query Match	Description
1	1474	10.3	328 9 US-09-925-297-816
2	982.5	6.9	238 9 US-09-729-835-80
3	430.5	3.0	3664 15 US-10-177-293-423
4	418	2.9	4019 10 US-09-738-973-425
5	418	2.9	4019 10 US-09-854-133-425
6	418	2.9	4019 15 US-10-144-649A-425
7	416.5	2.9	2665 9 US-09-864-761-34248
8	403	2.8	3256 10 US-09-919-172-98
9	403	2.8	3256 11 US-09-919-039-21
10	401.5	2.8	1367 10 US-09-801-368-108
11	401	2.8	2344 9 US-09-815-242-12713
12	391	2.7	5179 9 US-09-922-217-1068
13	391	2.7	5179 10 US-09-833-263-1068
14	391	2.7	5179 14 US-10-025-380-1068
15	384.5	2.7	3899 15 US-10-171-311-4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

THIS PAGE BLANK (USPTO)

Db 1120 SATSA-----SAGGTATPGKPPA--VVSQAAGS--TTVCATLTSVSTTSPSTA 1168
QY 1547 DKVKLMKFSRPKKTRSGTALPSYKFTVKTKSIFVLPNDLKKLARKGGIREVPYFNY 1606
Db 1169 SOLSI-----QLSSSTSTPLAETVVVSAH----- 1193
QY 1607 NAKPALDIWPPSPRPRTFGIWRYLQTVKSL--AGVSLMLRLWLASLRWDDMAKVPPG 1664
Db 1194 ----SLDKTSHSS--TTGLAFSLAPSSSSSPGAGVSYIS-----QPG 1231
QY 1665 GGSTRTFSEYETITTEIIKRRDVGPIGIREYCIKLIICPIGVPEPKETP--TPQRKGL 1723
Db 1232 G-----LHPLVIPSVIASPILPQAAGP 1254
QY 1724 RSSALRPKRPEPKOTGPV-----IETWVAEE-----ELE----- 1754
Db 1255 TSTPLLPQVPSIPPLVQVAVNPVAVQOTLIHSQPQALLPNQPHTHCPEVDSDTQPKAPG 1314
QY 1755 LWEIRAFARVE---KEKAQAVEQAKKRLQO---KPTV---IATSTTSPT---SSTTS 1802
Db 1315 IDDIKTLEEKLSRFSHSSGAQHASVLSLETSLVISTVTGPIPTTAVAPSKLLTSTS 1374
QY 1803 TISPAQKV---WVAPISGSVTTGKMWLTTKVGSPPATVTFQKNKNFHQTFTATVYKQGS 1858
Db 1375 TCLPPTNLPLGTVALPVPVTPG-----QVSTPVSTT----- 1407
QY 1859 NSGVVOVQOVKVLGIIPSSGTSGQOTFTSQPTATVTRPNISGGGTTNSQVITGPQI 1918
Db 1408 -----TSGVKPGTAPSKPPLTKAPVLPVGTGTELPAGTLPSEQLPFPF-- 1448
QY 1919 RQGMTVIRTPLOQ-----STLGRAIIRTPVMVOPGA--POOVMTQIIR 1959
Db 1449 GPSLTOSQOQPLEDDAQLRRTLSPEXITVTSVAVGVSMAAPTATEAGTQPKGVSOYKE 1508
QY 1960 GQPVSTAVSA-----PNTVSSSTPGQKS-----LTSATSTSNIOSSASQPP 1999
Db 1509 GPVLATSSGAGVFKMGFRQVSAADGAQKGNKSEDAKSVHFESSTSESSVLSSSS--- 1565
QY 2000 RPOQGVKLTMAQLTOLTQGHGNGQLTVVIOGOGOTTQOLOLIPQVTVLPQGGOLMQ 2059
Db 1566 -PESTLVKPE-----PNGITI--PGISSDVPE 1589
QY 2060 AAMPNGTVQRFLFTPLATTATVATTT-----TTVSTTAAGTGEQRQSK----- 2103
Db 1590 SAH-----KTTASEAKSDTQPTKVGRFQVTTTANKVGRFSVSKTEDKITDT 1636
QY 2104 -----LSPQMQRVHQDKTLPPAQS SVGPAKAPQTAQPSARPQOTQPO----- 2147
Db 1637 KKEGVPVAPPFMDLEQAVLP-----AVIPKKEKPELSEPHLNGPSPDPEAAFLSRDVED 1691
QY 2148 ---SPAQP-----PG----- 2152
Db 1692 GSGSPHSPHOLSKSLPQNLSQLNSLNSFNSSYMSNSDNEDEDLKLELRRLRDKHLK 1751
QY 2153 EVQ-----TOPEVQOTQTVSSHVP-----SEAQPTHQSSKQPQVAA---QSQ 2191
Db 1752 EIQDLQSRQKHEIESLYTKLGKVPVAVIIPPAAPLSGRRRRPTKSGKSSRSSSLGNKS 1811
QY 2192 PQ--SNVQGSFVRVQSPQTRIRPTSPQLSPQGSQOVQTTTSPQIPHTSLOIPSQ 2249
Db 1812 POLSGNLSGQSAASVHLPOQT--LHP-----PG-----NIPES 1842
QY 2250 GQPO--SOPQVS--STOTLSSGQTLN--QVSVSSPSRPOLOIQOP 2289
Db 1843 GQNLQLPLKPSPSDDNLYSAFTSDGAISVPSLSAPGQGIKQP 1885

Qy	2091	TTAAGTGEQRQSKLSP---QMQVHDKTLPPAAOSSSVG-----PAKAQPQT-----AQ	21336
Db	1337	TSNGGTGGPEGGOAPPAGRPCETHQTSTGTMTMSVSYGALLPDATSSHRTVESGLEVA	1396
Qy	2136	PSARPOQOTQOSP-----AOPEVOTQP-EVOTOTTYSHHVPSEAQTTHAOSKPKOV	2186
Db	1397	PSVTPOAGTALLAPFPFORVCNSPPCETHETGTHATTATVSNMSSNQDPPPRAASDOGEV	1456
Qy	2187	AASQPOSNVGGSFVRVQSPO-----TRIRPSTP---SOLSPGQSQSVQTTTSQPIPI	2238
Db	1457	ESTQGDVNITSSAITTTVTSSTILTRAIVTWTOSTPVPGPSVPPEELQVSPGRQQLP-	1515
Qy	2239	QPHTSLIQSGOGPOSQPOQVOSSTPTLLSGDTLNQVSVSSPRPQ-----LOIQO	2288
Db	1516	-PROLLQASTALMGESAELVLSAQTELPAAVDLSSTGEPSSGESAGSAVVATVVVVQP	1574
Qy	2289	PQPOVIAPVLQQVQLSOIQSQW-----AAAAAAAAAAAAAEQAALA	1634
Db	1575	PPPTQSEVDQLSLPQELLMAEAQACTTLMVTGLTPPELAVTAATAAAAAAAAEQAALA	1634
Qy	2330	LPLPIQIQSSAVOTHQIQNVVTVQARSVQEQLORVOQLRQDQKKKQOOIEIKREHTLQ	2389
Db	1635	IQAVIQAQAQAVMTGEGPMDTSEAAATVTAQELGHL-SABGQEQGQATTIPILTQOEELAA	1693
Qy	2390	ASNQSEIIQKVYMKHNNAVIEHLKQKKSMTPAE	2422
Db	1694	LVCQQQLQEAQAQOOH----HHLPTALAPAD	1721

RESULT 15

```

US-09-854-856-64
; Sequence 64, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 1911
; TYPE: PRN
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1911)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-64
    
```

QY	576	KLQEGKYRVYHNOYSTNSFALNKHQHRED-----HDKRRHLAHK-----FCLTPA	621
DB	242	ELMTSGTLKTYLKRFXVMKIKVLRSMCROQLKGLQFLHTPTPIIHRDLKCDNIFITGPT	301
QY	622	GEPKWNGSVHCSKVLTISTLTRTITTOLENNIPSSF LHPNWASHRAWIKAV-----	672
DB	302	GSVKIGD-----LGJATLK-----RASPASVIGTFEPMAP	332
QY	673	QMSCKPRE-----PALALILECA-----VKPY-----VMLPIW	701
DB	333	EMXEYEKDVESVDVAFGMCHLEMATSEYPYSECNAQAIIYRRVTSGVKPASFDPKVAIPEV	392
QY	702	RE-----FLGHTRLUHRMTSI-----EREKEK-----VKKK	727
DB	393	KEITEGCTRONKDERSYGIKOLLNHAFFOETGVRAELAEDDGECAIAKLWLRIEDIKKL	452
QY	728	EKKOEBEETMQOATWVXYTPPVKHQVWKQKEGYRYTG-----GGWSHIKTHYRFVPKL	784
DB	453	KGKYKDNEA-----IEFSFDLERDVEDVAQEMVESGYCEGDGHKTMAKA-----	497
QY	785	PGNTNVNRKSLEGTKNMNDENMDESOKRSRPKKIKTEPDSEKDEVKGSDAAKGADQ	844
DB	498	-----IKDRVSLIKRKREQOLVREEQEK-----KQEESLSKOOVQSASQ-----	540
QY	845	NEMDISKITEKODVVKELLDSOSDXPKCEEPMEVDDDMKTESHVNQCQESSQYDVWNVSE	904
DB	541	-----TGIKOLPSASTGIP-----TASTTSASVSTOVEPEBEPA	574
QY	905	GFHLRTSYKKTKSKLDGLLERIKOFTLEEKORLEKIKLEGKIGIKTSTNSSKNLS	964
DB	575	DQHOOLOYQPSTISVLSDGTVD-----SQGSSSVFT	605
QY	965	ESPVIITRAKEGCQSDSMRQEQSPANNNDOPEDLIQGCSQSDSVLRMSDSPSHTTNKLYPK	1024
DB	606	ESRV-----SSQOTVSYSQHEQAHSCTGTVPGHIPSTIVQAQSQP-----HGYYPP	650
QY	1025	DRVLLDVDVISRPFETKCPQNSIENDI-----EEKVSDLASRGQEP-----KSKT	1069
DB	651	SSVOQGI-----QOTAPPQOTVOYLSLQTSSTSEATTQAQVSPQAPQVLPVQSACKQST	705
QY	1070	KGNDFFIDDKSLASADDIGITLICKNKKPLLOESDTIVSSSKSALIHSSVPKSTINDRDATP	1129
DB	706	QG-----VSQVAPAEPAV-----AQP-QATQPTTLASSVSDSAHSDVASGMSGDNENV	752
QY	1130	LSRAMDFEGK-----LGCDSSENSTLENSSDT-----VS	1158
DB	753	PSSGRHEGRTTKRHYSVRSRSRHEKTSRPKLRLNVSNKGDRVVECCOLETHNRKMVT	812
QY	1159	IQ-----DSSEE-----DMIVONSNESISEOFRTREODVEVLEPLKCYLVSGSGTGN	1205
DB	813	FKFDLDGDNPEIATIMVNNDFILAIERESFVDQVR---ETIEKADEMLSEDYSVEPEGD	869
QY	1206	CEDRLPVKGTFAANGKPSQOKKLEER---PVNKCSDQIKLNKTTDKNNENRESEKKQR	1262
DB	870	-QGLESLOGKDDYGFGSCQKLEGEFKQPIPASSMPQQIGIPT-----	910
QY	1263	TSTFOINGKDNKPITYLKGE--CLKETSESRYVSGNVEPKVNNINKIIPENDIKSLTVKE	1320
DB	911	SSLTQV-----VHAGNRRFIVSPPERLRESKVFP-----SEITDIVAAS	951
QY	1321	SAIRPFTGNDVIMEDFNERNSETKSHLLSSDAEG-----NYRDSLETLPSTK-----ES	1371
DB	952	TAQSPGMN---LSSHASSLSLQQAFAFSELRLRAQMTEGPNAPTAPNFSGTGFVVVPPPLSS	1008
QY	1372	DSTOTTTPPASCPSPSNVNQVEDMEIETSEVKVYTSPTISEESNL-----SNDFTDEN	1426
DB	1009	IAGVPTTAAATAAPYPATSSPND--ISTS---VIQSEVTVPTBEGIAGVATSTGVVTS	1062
QY	1427	GLPKNNKNVNGESKRKTIVTEVTMTSTVATESKTIKVEKGDKQOVVSTENCASKSTV	1486
DB	1063	GLPI---PPYESPVLSSVSSIITPAVVSISTTSPSLQWTPTSTSELVSWSTALYPSVTV	1119
QY	1487	TTTTTTVTKLTSTPSTGGSGVDIIISVKEOSKTVTTTVDTSLTGTTGTLVTNTVTSKEYSTR	1546

1107 -----PPCETHETGTTTATAMSSVGNHQRARRACAACTPAVIRISVATGAL 1156
1974 SSTPGOKS-----LTSATSTSN-----QSSASQPP-----RPOQG 2004
1157 EAAQGSQOCOTROTSATSTTTVMATGAPCSAGPLLGPSMAREPGRSAPFVQLAPLSS 1216
2005 QVKLTMAQLTQTQHGNGOGLTVVIOGOGTTGLOLIP-----QG-----VTVLP 2051
1217 KVRLLSPSFKDLPAGRHSHAVSTAAMTRSSVAGPRMAPVCESLQGGSPSTVTTVALE 1276
2052 G---PQOQLMO-----AAMPN-GTVORFLETPPLATTATTASTTTTTS 2090
1277 ALLCPSATVTQVCSNPPCETHETGTTNTATTNAGSAQRVCNPPCETHETGTTTATTA 1336
2091 TTAAGTGEOROSKSLP---OMVHQDKTLPPAQSSVG-----PAKAQPT-----AQ 2135
1337 TSNGGTGQEGGQPPAGRPCETHQTSTGTTMSVSGALLPDATSSHRTVBSGLEVA 1396
2136 PSARPQOTQOSP-----AQPEVOTQP--EVOTQTTVSSHVPSEAOPTHAQSKPKOV 2186
1397 PSVTPQAGTALLAPFTQRCVSNPPCETHETGTTTATTVTNMSNQDPPPAASDOGEV 1456
2187 AAQSPQSNVQGSVRVQSPSQ-----TRIRPSTP---SOLSPGQOSQVQTTTSQPIPI 2238
1457 ESTQGSVNITSSAITTTTSSVSTLTRAIVTTQSTPVPGPSVPPPEELQVSPGPRQQLP- 1515
2239 QPHTSLQIPSGQPOQPOQVQSTQTLSSQOTLNOVSVSPSRPQ-----LQIOQ 2288
1516 -PRQLQASATLGMESAELVASQTPPELPAADVLSLSTGEPSSGQBSAGSNAVTVVQV 1574
2289 PQQVIAPVLOQOQOVVLSQIQSOV-----AQIQAQSGVGPQIQK 2329
1575 PPPTQSEVDLSLPQELMAEAQAGTTLMVTGLTPELATVAAEAQAQAAAEQAALA 1634
2330 LQLPIQIQSSAVOTHOIQNVTVVQAASVQEQIORVQOQLRDQOQKKQOQIEKREHTLQ 2389
1635 IQAVLQAQAAQVMTGPEPMDTSEAAATVQAELGHL-SARGOQGOATTIPIVLTOQELAA 1693
2390 ASQSELIQOVVKNHNAVIEHLKOKKSMTPAE 2422
1694 LVQOQOQOEAQAQOQH-----HHLPTALAPAD 1721

RESULT 14
PCT-US93-11721-5
; Sequence 5, Application PC/TUS9311721
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11721
; FILING DATE: 03-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-57503-1/RAO

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-11721-5

Query Match 2.5%; Score 353; DB 5; Length 2035;
Best Local Similarity 21.4%; Pred. No. 2.1e-12;
Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;

QY 1368 TKESSTOTTT-PSASCPESSNVQVEDMETSEVKKVT-----SSPITSEESNLSN 1420
DB 619 TSVSSATNTSPRIITVHKSGTVAQAQVVTTVGGVTITITLVKSPISVPGGSAL--- 676
QY 1421 DFIDENGLPINKNNVNGESKRKTVITEVTMTSTVATESKT-----VIKVE---KGDK 1471
DB 677 -----ISNLGKVMNVQKPKVQTSVATGQASTGPTVQIIQTKGPLPAGTI 721
QY 1472 QTVVSTENCASKTVTTTFTVKLSTPSTGSGVDIIISVKEQSKTVVTTTDS----- 1525
DB 722 LKLVTSAD--GKPTTIIITTTQASGAGTKPT--ILGISSVSPSTTKPTTIIKTIPMSAI 777
QY 1526 LTTTGCTLVSTMTVSKYSTRDVKMLKFSRPKTRSGTALPSY-----RKFTVKSTKS 1580
DB 778 ITQAGATGVTSSPGTKSPITITTKVMTSGTGAPAKIITAVPKIATGHQCGQGTQVVLKG 837
QY 1581 IFVLPNDLKLARKGGIREVPVFNNA-KPALDIMPYSPRPTGTTWRYLQTVKSLA 1639
DB 838 APQOPGTLRTVP-MGVRILVPTVTVSAVKPAVTLVW---KGTGVT---TLGTVTGV 890
QY 1640 GVSMLRLLWASLRWDDMAAKVPPGGGSTRTESEITEITTEIKRRDVGPGYIREYCI 1699
DB 891 STSL-----AGAGHSTASLATPITTLGTIATLS----- 920
QY 1700 RKIICPIGVPEPKETPTQKGLRSSALRPKRPTPKQ---TGPVITETWVAEEL 1755
DB 921 SQVINPAINVSAQTLTAAGGLTPTITMOPVSQPTQVTLITAFSGVE----- 970
QY 1756 WEIRAFAEERVEKEKAQAVEQOAKRLEQOKPTVIATSTSPSTSTSTISPAQKVMVAPI 1815
DB 971 -----AQPV-----HDLVPSILASPTTEQPTATVTIADSGQDVQP- 1006
QY 1816 SGSVT-----TGTKMWLTTKV-----GSPATVTFQONKFNHOTFATWVKQ--G 1856
DB 1007 -GVTLVCSNPPCETHETGTTNTATTTVVANLGHGHPQTPQVQVCDRQEAASLIVTSTVG 1065
QY 1857 QNSGVVQVQOQKVLGIIPSSSTGTSQOTFTSFQPTATVTIRPNTSGSGGTTSSQVITGP 1916
DB 1066 QQNGSVVRVCNPN---PCET---HETGTTNTATTAT---SNMAGHGCN----- 1106
QY 1917 QIRPGMTVIRTPLOQSTLGAIIITPVWVOPGAPOQ--VMTQIIRGQPVSTAVS-APNTV 1973
DB 1107 -----PPCETHETGTTNTATTAMSSVGNHQRARRACAACTPAVIRISVATGAL 1156
QY 1974 SSTPGOKS-----LTSATSTSN-----QSSASQPP-----RPOQG 2004
DB 1157 EAAQGSQOCOTROTSATSTTTVMATGAPCSAGPLLGPSMAREPGRSAPFVQLAPLSS 1216
2005 QVKLTMAQLTQTQHGNGOGLTVVIOGOGTTGLOLIP-----QG-----VTVLP 2051
1217 KVRLLSPSFKDLPAGRHSHAVSTAAMTRSSVAGPRMAPVCESLQGGSPSTVTTVALE 1276
2052 G---PQOQLMO-----AAMPN-GTVORFLETPPLATTATTASTTTTTS 2090
1277 ALLCPSATVTQVCSNPPCETHETGTTNTATTNAGSAQRVCNPPCETHETGTTTATTA 1336

```
Db 971 -----AQPV-----HDLVSVILASPTTEQPTATVTIADSGGDVQP- 1006
Qy 1816 SGSVT-----TGKMWLTTKV-----GSPATVTFQKNKHFQTFATWKVQ--G 1856
Db 1007 -GTVTLVCSNPPCETHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQEAASLVTSTVG 1065
Qy 1857 QNSGVVQVQKVLGIIPSSGTSTGTSTQPTATVTIRPNTSGSGGTTNSQVITGP 1916
Db 1066 QONGSVVRVCSNP-----PCET-----HETGTTNTATTAT-----SNMAGQHGCN- 1106
Qy 1917 QIRPGMTVIRTPLOQSTLGKAIIRTPVMVQVQAPQ--VMTQIIRGQPVSTAVS-APNTV 1973
Db 1107 -----PPCETHETGTTNTATTAMSSVGANHQDRACCAAGTPAVIRISVATGAL 1156
Qy 1974 SSTPGQKS-----LTSATSTSN-----QSSASOPP-----RPOQG 2004
Db 1157 EAAQGSKSOCQTRQTSATSTMTVMATGAPCSAGPLLGPMAREPGRSFAFVLQAPLSS 1216
Qy 2005 QVKLTMAQLTQTOGHGNGQLTVVIOGQGTTCQLQIP-----QG-----VTVLP 2051
Db 1217 KVRLLSPSIKDLPAGRHSHAVSTAAMTRSSVSGAGEPRMAPVCELSQGGSPSTTVTVALE 1276
Qy 2052 G---PGQQLMQ-----AAMPN-GTVQRFLETPLATATTASTTTTTSV 2090
Db 1277 ALLCPSATVTVQCSNPPCETHETGTTNTATTSNAGSAQRVCSNPPCETHETGTTHTTATTA 1336
Qy 2091 TTAAGTGEQRQSKLSP---QMQRHDKTLPPAQSSVG-----PAAQAPQT-----AQ 2135
Db 1337 TSNGGTGQEGGQPPAGRPCETHQTTSTGTTMSVSVGALLPDATSSHRTVESGLEVAAA 1396
Qy 2136 PSARPQPTQPOSP-----AQEVQTOP--EVOTQTTVSHVPSAQPTHAOSSKPQV 2186
Db 1397 PSVTPQAGTALLAPFPTRQVCSNPPCETHETGTTHTTATTVTSNMSNQDPPPAASDQGEV 1456
Qy 2187 AAQSQPOSNVQGSQPVVRQSPQ-----TRRPSTP---SOLSPQOSQVQTTTSOPIPI 2238
Db 1457 ESTOGDSVNTSSAIIITVSSITLRAVTVTQSTPVPGRVSPPELQVSPGRQOQLP- 1515
Qy 2239 QPHTSLOIPSGQPOSQPOVQOSSQTLSGGQTLNOVSVSPSRPQ-----LQIQQ 2288
Db 1516 -PROLLQASASTALMGESAELVLSAQTPELPAADVLSSTGSPSSQESAGSAAVAVTVVQP 1574
Qy 2289 POQVIAVPLOQQOVQVLSQIQSOVV-----AQIAQOQSGVPOQIK 2329
Db 1575 PPTQSEVDQSLPQELMELAAQAGTTLMVTGLTPEELAVTAAEAQAAATBEAQALA 1634
Qy 2330 LQLPIQIQQSAVOTHQIQNVTVVQAASVQEQVQOLRQDOQKQKQOQLEIKREHTLQ 2389
Db 1635 IQAVLQAAQAAVMTGEPMDTSEAAAVTQAEGLHL-SAEQEQEQATTIPIVLTOQEELAA 1693
Qy 2390 ASNOSEIIQKQVNMKNHNAVIEHLKQKKSMTPAE 2422
Db 1694 LVQOQOQLQEAQAQOQH-----HDLPTALAPAD 1721
```

RESULT 13

US-08-393-703-5

; Sequence 5, Application US/08393703

; Patent No. 5585239

; GENERAL INFORMATION:

; APPLICANT: Lamarco, Kelly

; APPLICANT: Wilson, Angus

; APPLICANT: Herr, Winship

; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:

; NUMBER OF INVENTION: HOST CELL FACTOR

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

```
2IP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,703
FILING DATE: 24-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-2/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-393-703-5
```

Query Match 2.5%; Score 353; DB 1; Length 2035;

Best Local Similarity 21.4%; Pred. No. 2,le-12;

Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;

```
Qy 1368 TKESDSTQTTT-PSACPCESNSNOVEDMEIETSEVKKVT-----SSPITSEESNSLN 1420
Db 619 TSVSSATNTRPITVHKSGTVTAQQAQVTVVGVGTVTKTITLVKSPISVPVGSAL-- 676
Qy 1421 DFIDENGLPINKNNVNGESKRKTVITETVTMTSTVATESKT-----VIKVE---KGDK 1471
Db 677 -----ISNLGKVMVSVQTKPVQTSVTCQASTGPVTVQIIQKGPLPAGTI 721
Qy 1472 QTVVSSSTENCASKSTVTTTITTTTSTPSTGSGVDIISVKBQSKTVTTTVDTS----- 1525
Db 722 LKLVTSAD--GKPTTITTTQASGAGTKPT--ILGISSVSPSTTKPGTTTIKIPTMSAI 777
Qy 1526 LTTTGGTLVTSMTVSKYSTRDKVKLMKFSRPPKTRSGTALPSY-----RKFTVSKTKKS 1580
Db 778 ITOAGATGVTSSPGIKSPITITTKVMTSGTGAPAKIITAVPKIATGCGQGVQVWLKG 837
Qy 1581 IFVLPNDDLKLRKGGIREVPYFNNA-KPALDIWVPYSPRPTFGITWRYRLQTVKSLA 1639
Db 838 APGQPGTILRTVP-MGVRLVTPVTVSAKPAVTVLV---KGTGVT---TLGTVTGTV 890
Qy 1640 GVSMLRLWLASLRWDDMAAKVPPGGGSTRTSETETITTEIIRKRDVGPYIGIRFEYCI 1699
Db 891 STSL-----AGAGGHSTASLATPITTLGTIATLS----- 920
Qy 1700 RKIICPIGVPTPKETPTPQRKGLRSSALRPKRETPKQ-----TCPVLIETWVAEELEL 1755
Db 921 SOVINPTAITVSAQATLTAAAGLTTTITMQPVSOPTQVTLITAPSGVE----- 970
Qy 1756 WEIRAFARVEKEKAQAEQAKKRLQEQKPTVIATSTTSPSTSTTSPISPAQVMVAPI 1815
Db 971 -----AQPV-----HDLVSVILASPTTEQPTATVTIADSGGDVQP- 1006
Qy 1816 SGSVT-----TGKMWLTTKV-----GSPATVTFQKNKHFQTFATWKVQ--G 1856
Db 1007 -GTVTLVCSNPPCETHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQEAASLVTSTVG 1065
Qy 1857 QNSGVVQVQKVLGIIPSSGTSTGTSTQPTATVTIRPNTSGSGGTTNSQVITGP 1916
Db 1066 QONGSVVRVCSNP-----PCET-----HETGTTNTATTAT-----SNMAGQHGCN- 1106
Qy 1917 QIRPGMTVIRTPLOQSTLGKAIIRTPVMVQVQAPQ--VMTQIIRGQPVSTAVS-APNTV 1973
```

QY 1487 TTTTITVTKLSTPSTGSGVDIISVKEQSKTVVTTTITVDTSLTTGGTGLVMTSVKEYSTR 1546
Db 1180 SATSA-----SAGSRTATPGKPPA--VVSQQAAGS--TTVGATLTSVSTTTSPSTA 1228
QY 1547 DKVKLKFRRPKKTRSGTALPKYKRVKTSKKSIFVLPNDLKLARKGGIREVPFNY 1606
Db 1229 SQLSI-----QLSSSTSTPLAETVVSAN----- 1253
QY 1607 NAKPALDIPYSPRPTFGITWRYRLQTVKSL--AGVSLMLRLLWASLRWDDMAAKVPPG 1664
Db 1254 ----SLDKTSHS---TTGLAFSLSPSSSSPGAGVSYIS-----QPG 1291
QY 1665 GGSTRFETSETITTEILIKRVDGVPYGIREFEVCIRKIICPGVPETPKETP--TPQRKGL 1723
Db 1292 G-----LHPLVPSVIASTPLPQAAQP 1314
QY 1724 RSSALRPKRPETPKQGPV-----IETWVAEE-----ELE----- 1754
Db 1315 TSTPLLPQVPSIPPLVOPVANVPVAVQOTLIHSQOPALLPNQPHTHCPEVDSDTQPKAPG 1374
QY 1755 LWEIRAFARVE---KEKAQAVEQAKKLEQO---KPIV---IATSTTSPT---SSTWS 1802
Db 1375 IDIKTLEKRLSLFSEHSSGAQHASVLSLETSLVETVPGIPTAVAPSKLLTSTS 1434
QY 1803 TISPAQKV----MVAPTSQVTTGTAKMLTKVGSPTVTFQONKFNHQTATWVRQOGS 1858
Db 1435 TCLPPTNLPGTVALPVPVTPG-----QVSTPVST----- 1467
QY 1859 NSGVQVQVKGLIIPSSGTGSOOTFTSFOPRATVIRPNTSGSGGTSTNSQVITGPQI 1918
Db 1468 -----TSQVKGTAPSKPPLKAPVLPVGTLPAGTLPSEQLPFPF-- 1508
QY 1919 RGMVTVRTPLQO-----STLGKALIRTPVMVQPCA--PQOVMTQIIR 1959
Db 1509 GPSLTOSQQLLEDLDAQLRRLTSPEXITVTSVAVGPVSMAPTAITEAGTQPGVSOVKE 1568
QY 1960 QGPVSTAVSA-----PNTVSTTPQKS-----LTSATSTSNIOSSASOPP 1999
Db 1569 GPVLATSSGAGVFKMGFRQVSVAAQAGKKNKSEDAKSVHFESSTESSVLSSSS-- 1625
QY 2000 RPOGQVKLTMAQLTQLTOGHGNGQLTVVIOGQGTGQLQLIPQGVTVLPQGOQLMQ 2059
Db 1626 -PESTLVKPE-----AVIPKKEKPELSEPHLNGPSSDPEAAFLSRDVED 1751
QY 2060 AAMPNGVQVRFLETTPLATTATTTT-----TTVSTTAAGTGBQROSK----- 2103
Db 1650 SAH-----KTTASEAKSDTGPQTKVGRFQVTTANKVGRFSVSKEDKITDT 1696
QY 2104 -----LSPQMVHQDKTLPPAOSSSVGPAAQPTAQAQSAQPOQTPQO----- 2147
Db 1697 KKEGPVAPPPFMDLEQAVLP-----AVIPKKEKPELSEPHLNGPSSDPEAAFLSRDVED 1751
QY 2148 ---SPAOP----- 2152
Db 1752 GSGSPHSPHOLSSKSLPSQNLSSLSNSFNSSNSYMSDNEIDEDLKLRLRLDKHLK 1811
QY 2153 EVQ-----TOPEVQOTQTVSSHVP-----SEAQPTHAQSSKQVAA---QSQ 2191
Db 1812 EIQDLQSRQKHEIESLYKLKGVPVAVIIPPAPLSGRRRRPTKSKGSSRSSLGNKS 1871
QY 2192 PQ---SNVQGSVPRVQSPQTRIRPSTPSQLSPGQSQVQVQTSQPIPIQPH----- 2242
Db 1872 POLSGLNLGGQSAASVLHPQOT--LHP--PGNIPESGNQL-----LQPLKPSPSDNLXSAF 1924
QY 2243 ----SLOIPSGQPOQPOVQSSQTSOTLSSGOTLNQVSVSPSRPOLQIQPOQ 2291
Db 1925 TSDGAISVPSLSAP---GQGTSTSTNTV--GATVNS-----QAAQAP 1961

RESULT 12

US-08-046-585-5

; Sequence 5, Application US/08046585

; Patent No. 5453362

GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Willson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/046,585
; FILING DATE: 12-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-046-585-5

Query Match 2.5%; Score 353; DB 1; Length 2035;
Best Local Similarity 21.4%; Pred No. 2,1e-12;
Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;
QY 1368 TKESDSTQTTT-PSACSPESVNVQVEDMEIETSEVKVVT-----SSPTSEESMLSN 1420
Db 619 TSVSSAINTSTRPIITVHKSGTVTAQQAQVWTVVGGVTKTITLVKSPISVPGGSAL-- 676
QY 1421 DFIDENGLPINKNENGVESKRTVITEVITMTSTVATESKT-----VIKVE---KGDK 1471
Db 677 -----ISNLGKVMVSVQTKPVQTSVAVTQASTGPVTOIOTKGPLPAGTI 721
QY 1472 QTVVSSSTENCAKSNVTTTTTNTVTKLSTPSTGSGVDIISVKEQSKTVVTTTVD----- 1525
Db 722 LKLWTSAD--GKPTIITTTQASGAGTKPT--ILGIVSVSPSTTKPGTTIITKIPMSAI 777
QY 1526 LTTTGGTGLVSMTVSKSEYSTRDKVKMLKFRPKKTRSGTALPSY-----RKFTVKTSKS 1580
Db 778 ITQAGATGVTSPPGKSPITITTKVMTSGTGAPAKIITAVPKIATGCGQGVTVVYLGK 837
QY 1581 IFVLPNDDLLKARKKGIRVFPYFNNA--KPALEDIWPYSPRPTFGITWRYRLQTVKSLA 1639
Db 838 APQGPGLITRTP--MGVRLVTPVTVSAVAPVTLV---KGTGTV---TLGTVTVGTV 890
QY 1640 GVSMLRLLWASLRWDDMAAKVPPGGSTRTTETSETITTEIIRKRDVGPYGIREFYCI 1699
Db 891 STSL-----ACAGHSTASLATPITLTGATLIS----- 920
QY 1700 RKIICPIGVPTPKETPTPQKRLSRSSALRPKRPETPKQ---TGPVITWVAEELEL 1755
Db 921 SQVINPTAITVSAOITTLTAAGGLTTPTITMQPVSQPTQVTLITAPSGVE----- 970
QY 1756 WEIRAFARVEKEKAQAVEQAKKRLQEQKPTVATSTTSPTSTTSTISPAQVMVAPI 1815

Qy	2000	RPQGVQVKLFWAQITQLTOGHGNGQLTVVIOGQQTGOLQILPOGVTVLPGPQOQLMQ	2059
Db	1566	-----PSTLVKPE-----PNCITIL-PGSISSDVPE	1589
Qy	2060	AAMPNGCTVQRELFPLATTATTASTTT-----TTVSTTAAGTGEOROSK-----	2103
Db	1590	SAH-----KTTASEAKSDIGQPTKVGRFQVTTTANKVGRFSVSKTEDKITDT	1636
Qy	2104	-----LSPQMQVHQDKTLPPAQQSSVGPAPAKAQTPAQSARPOPTQPO-----	2147
Db	1637	KKEGPVASPFMDLEQAVLP-----AVIPKKEPSELSEPHLNGPSSDPEAFLSRDVDD	1691
Qy	2148	-----SPAQP-----	2152
Db	1692	GGSPHSPHOLSKSLPSQNLSSLSNFSNSYMSDNESDIEDLKLRLRLRDKHLK	1751
Qy	2153	EVO-----TOPEVQOTQTVSSHVP-----SEAQPHQAOSSKPQVAA---QSQ	2191
Db	1752	EIODLSQRKHIEHSLYTKLGKVPVAVIIPPAAPLSGRRRRPTKSGKSSRSSLSGNKS	1811
Qy	2192	PQ---SNVQGSQPVRYOSPSQTRIRPSTPSQLSPQQSQVQTTSQTPIPIQPHT-----	2242
Db	1812	PQLSGNLSGGQSAASVHLHPQT-LHP--PGNIPESGQNL---LQPLKPSPSDNLYSAF	1864
Qy	2243	-----SLQIPSGQGPQSQPVQOSTQTQLSSCQTLNQVSVSPSRPQIQIQPQP	2291
Db	1865	TSDBGATVPSLSAP---GQCTSSNTV--GATVNS-----QAAQAQP	1901

```

RESULT 11
US-09-854-856-30
; Sequence 30, Application US/09854856
; Patent NO. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854, 856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 2108
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2108)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-30

```

	Query Match.	2.5%;	Score 358.5;	DB 4;	Length 2108;
	Best Local Similarity 18.1%;		Pred. No. 1.1e-12;		
	Matches 411;	Conservative 295;	Mismatches 716;	Indels 851;	Gaps 96;
Qy	399	TNARGSNKSFLEAAANEELLESIRAKGDDIDNVKSPETEEDKDNFENDSKDAEKNEEF	458		
Db	160	TSKDRPVQSBLVGSKEEPPARSGSGG	-----SAKEPQERSQQODDI-----EEL	207	
Qy	459	EDQSLKSDKTPDDPEQCKSEVGFDFKSEKSNGLSESPGAKGASGSTRI-ITRLRN	517		
Db	208	ETRAVGNMNDGRFLKFDIEIGR	--GSFKT-----VYKGLDTEITTEVAVANCEL	252	
Qy	518	PDSKLSQLKSQQVAAAAAHEANKLFKEKEKVLVYNSQGEISRLSTKTKKIVIMKGN-	INNYF	575	
Db	253	QDRLLTKYKSERQ	-----FKGEAEMLKGLQHPNIVRFYDSWESTVYKGGKCIVLVT	301	

Qy	576	KLQGEGRYVYHNOYSTNSFALNKHQRED-----HDKRRHLAHK-----FCLTPA	621
Db	302	ELMTSGTLKTVLKREKVMKIKVLRWCQRILKGLQFLHTRTPPIIHRDLKCDNFIINGPT	361
Qy	622	GEFKWNGSVHGSKVLTITSTLRITTOLENNIPSSFLPNWASHRANIKAV-----	672
Db	362	GSVKIGD-----LGLATLK-----RASPAKSVIGTPEFMAP	392
Qy	673	QMSCKPRE-----FALALATILECA-----VKPV-----VMLPIW	701
Db	393	EMYEKKYDESVDYAFGCMHLEMATSEYPYSECONAAQIYRRVTSYGVKPAFDKVAIPEV	452
Qy	702	RE-----FLGTHRLHRTSI-----EREKEK-----VKKK	727
Db	453	KEELGECIRONKDERYSIKDLLNHAPFOETGVRVELAEBDDGKIAIKLWLRIEDIKKL	512
Qy	728	EKKQEBEETWQATWVYKTPVKHQWKQKGEERYVTGY---GGWSWKSTHYVRFVKL	784
Db	513	KGKYKDNEA-----IEFSFDLDRDVPEDAQEMVESGYCEGDHKTMAKA-----	557
Qy	785	PGNTNVNRYKSLGTTKNMNDENDESDKRCSRPPKKIKTEPQSEKDEVGSDAAKGAQD	844
Db	558	-----IKDRVSLIKRKREQRQOLVREBEKK-----KQESSLKQOVQSSASQ-----	600
Qy	845	NEMDISKITEKKQDOVKELLDSDSDKPCKEPMEVDMDKTESHVNCQESSQYDVVNVSE	904
Db	601	-----TGIKQLPSASGIP-----TASITTSASVSTQVEPEPEA	634
Qy	905	GFHLRTYKKKTKSKLDGLLERIKOFTLEEKORLEKIKLEGKIGIGTKTNSNKNLS	964
Db	635	DQHOOLQYQOPSISVLSDGTVD-----SQGSSSVFT	665
Qy	965	ESPVITRAKEGCQSDSMRQEQSPNANDQPEDLLQGCQSQSDSVLVRMSDPSTHNTKLYPK	1024
Db	666	ESRV-----SSQQTVSYGQHEQAHSTGTVPGHIPSTVQAQSQP-----HGVIYP	710
Qy	1025	DRVLDDYSIRPETKCPKQNSIENDI-----EEKVSDLASRGOEPT-----KSKT	1069
Db	711	SSVQGI-----QQTAPQQTQVYLSLQSTSTSEATTQAQVSPQAPQVLPQVYSAGKQST	765
Qy	1070	KGNDFFIDDKSLASADDIGTLICKNNKPLQEBSDTTVSSSKSALHSSVPKSNDRDATP	1129
Db	766	QG-----VSQVAPAEPAV-----AQP-----QATOPTTLASSVDSAHSDVAGSMGDNENV	812
Qy	1130	LSRAMDFEGK-----LGCDSESNSTLENSSDT-----VS	1158
Db	813	PSSSGRHEGRTTKHYKKSVRSRHRBKTSRPKURLINLVNKKGDVVQCOLETHNRKMWI	872
Qy	1159	IQ-----DSSEE-----DMIVQNSNESISEQFRTREQDVEVLEPKLCELVSGESTGN	1205
Db	873	FKFDLDGDNPEEATINVNDFILAIRESFVDQVR---EIEEKADEMLSEDSVVEPEGD	929
Qy	1206	CEDRLPVGTEANGKPSQOKKLEER---PVNKGSDQIKLNTTDDKNNENRESEKKQGR	1262
Db	930	-QGLSESQGGDDYGFSGSOKLEGEFKQIPASSMPQOIGIPT-----	970
Qy	1263	TSTFOINGKONKPIYLLKGE--CLKEISESRVYSGNVPEKVNINKIIPENDIKSLTVKE	1320
Db	971	SSTQV-----VHSAGRFRFVSPPEKRLRESKVFP-----SEITDVTVAAS	1011
Qy	1321	SAIRPFINGDVIMEDFNERNSSETKSHLLSSDAEG-----NYRDSLETLPSTK-----ES	1371
Db	1012	TAQSPGN---LSHASSSLQQAFLSELRAQMTGEGTAPPNFSHGTGTFPVVPPLSS	1068
Qy	1372	DSTOTTTTPSACPSBNSVNOVEDMEIETSEVKVYKVTSSPITSEESNL-----SNDFIDEN	1426
Db	1069	IAGVPTTAAATAPVATSSPPND--ISTS---VIQSEVTVPTEEGIAGVATGTGVVTS	1122
Qy	1427	GLPINKNNENVGESKRKTVITEVTTMTSTVATESKTVIKVEKQDKQVWVSTENCAKSTV	1486
Db	1123	GLPT---PPVSESPVLSSVVSSTIIPAWVSITSTSPLOSPTSETVSEWSTALYPSVTV	1179

```

; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854, 856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 2048
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2048)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-62

Query Match          2.58; Score 358.5; DB 4; Length 2048;
Best Local Similarity 18.18; Pred. No. 1e-12;
Matches 411; Conservative 295; Mismatches 716; Indels 851; Gaps 96;

QY 399 TNKARGSNKSLAAANEIILESIRAKKGDIDNVKSPETEDKKNETENDSKDAEKNREEF 458
DB 100 TSKDRPVQSPLVSGSKEEPPARSGGG-----SAKEPQERSQQODDI-----EEL 147

QY 459 EDQSLKSDDKTDPDPEQKSEVGFKSEKSNGLSELSESPGAKGASGSTRI--ITLRN 517
DB 148 ETKAVGNSDGRFLKFDIEIGR--GSFKT-----VYKGLDTETIVEAVMCEL 192

QY 518 PDKLSQLKSOQVAAAHAEANKLFKEGKEVLVNVNSOGEISRLSTKKEVIMKGN--INNIF 575
DB 193 QDRMLTKSERQ-----EKERAMLKGLQHNIVRFPYDSWESTVKGKCCIVLVT 241

QY 576 KLQEGKYRVYHNOYSTNSFALNKHORED-----HDKRRHLAHK-----FCLTPA 621
DB 242 ELMTSGTLTKYLRKFKVMKIKVLRWCQILKQLGLFHLTRTPPIIHRDLKCDNIFITGPT 301

QY 622 GEFKNGSVHGSKVLITSTLTTLTQLENNIPSSFLHPNASHANWIKAV-----672
DB 302 GSVKIGD-----LGLATLK-----RASFAKSVIGTPEFMAP 332

QY 673 QMCKSPRE-----FALALAILCA-----VKPV-----VMLPIW 701
DB 333 EMYEKYDESVDVYAFGCMLEMATSEYPSQCQNAQIYRRVTSGVKPAFSDKVAIPEV 392

QY 702 RE-----FLGTRLRHMTSI-----EREKEK-----VKKK 727
DB 393 KEIEGCIQNKQDERYSIKDLLNHAFFQETGVRVELAEEDDGEKIAIKLMLRIEDIKKL 452

QY 728 EKKQEEETMQOATWVKYTFPVKHQVWKQKGEYRVGY-----GGWSWISKTHVYRFVKL 784
DB 453 KGKYKDNEA-----IEFSFDERDVPEDVAQEMVESYVCEGDHKTWAKA-----497

QY 785 PGNTNVNVRKSLGFTKNNNDENMDESKRCSRPKKIKIEPDSKEDEVKGSADAKGADQ 844
DB 498 -----IKDVRSLIKKRQORQLVREEQKK-----KQESSLKQOQVEQSSAQ-----540

QY 845 NEMDISKTEKKDQDVKELLDSDSKPKCEKPEMVEDDDMKFTESHVNCQESSQVDVVNYS 904
DB 541 -----TGIKOLPASTGIP-----TASTTSASVSTQVEPEPEA 574

QY 905 GFHLRTSYKKTKSKLDGILLRRIKQFTLEEKORLEKIKLGGIKGICKTSTNSKNLS 964
DB 575 DQHQOLOYQOQPSIVLSDGTVD-----SGQGSVFT 605

QY 965 ESPVITKAKEGQSDSMRQEQSPNANNQPEDLIQCCSQSDSVLRMDPSHTTNKLPK 1024
DB 606 ESRV-----SSQOTVSYGSOHAHQHSTGTVPGHIPSTVQAQSQP-----RGVYPP 650

QY 1025 DRVLDDVSIRSPTKCPKQNSIENDI-----EEKVDSLASRGQEP-----KSKT 1069
```

```

DB 651 SSVQOQI-----QQTAPPOQTVOYLSQTSSTSSSEATTQAQVPSQPAQVLPQVSAGKQST 705
QY 1070 KGNDFFIDDSKLASADDIGTLCKNKKPLIOESDITIVSSSKSALHSSVPKSTNDRDATP 1129
DB 706 QG-----VSQVAPBPVAV-----AQP--QATQPTTLASSVDVSAHSDVSGSDGNEV 752
QY 1130 LSRAMDFEGK-----LGCDSSENSTLENSSDT-----VS 1158
DB 753 PSSSGRHEGRTTKRHYRKSVRSRHEKTSRPKRLILNVSNKGRDVRVEQOLETHNRKMVT 812
QY 1159 IQ-----DSSEE-----DMIVQNSNISISQFRTREODVEVLEPLKCELVSGETGN 1205
DB 813 FKFDLDGDNPEELATIMVNNDFILAIERESFVDQVR---EIEEKADEMSEDSYSEPEGD 869
QY 1206 CEDRLPVKTEANGKKPSQOKLEER---PVNCKSQDIKLNKNTDKNNENRESEKKQOR 1262
DB 870 -QGLSLOGKDDYFGSGQKLEGEFKOPIPASSMPOQIGPT-----910
QY 1263 TSTFQINGKDNKPKIYLKGE--CLKEISESRVSGNVPEPKVNNINKIIPENDIKSLTVKE 1320
DB 911 SSLTQV-----VHSAGRRFIVSPVPSRLRESKVPF-----SEITDTVAAS 951
QY 1321 SATRPFINGDVIMEDFERNSETKSHLLSSDAEG-----NYRDSLETLPSK---ES 1371
DB 952 TAQSPGMN---LSHSASSLQQAFASELRAQMTGENTAPPNFSHTGPTFPVVPPLSS 1008
QY 1372 DSTQTTTPSACSPESNVNOVEDMEIETSEVKVTSPTSEESNL-----SNDFIDEN 1426
DB 1009 IAGVPTTAATAPVATSPSPND--ISTS---VIQSEVTPTEEGIAGVATSGVVTSG 1062
QY 1427 GLPINKNENNGESKRKTVITEVMTSTVATESKTVIKVEKDGKQTVVSTENCASSTV 1486
DB 1063 GLPI---PPVSESVLSVSVSSITIPAVVSIISTSPSLQVPTSTSELVSVSTALYPSVT 1119
QY 1487 TTTTITTKLSTPSTGSGVDIISVKEQSKTVTTTITVTDLSLTGGLVMTSVSKYESTR 1546
DB 1120 SATSA-----SAGSTATPGKPPA--VVSQQAAGS--TTVGATLSTSTTTSFSTA 1168
QY 1547 DKVKLMKFRPKTRSGTALPSYRKFTVTKTKKSFIVLPNDLKLARKKGIREVPFNY 1606
DB 1169 SLSI-----QLSSSTSTPLAETVVVSAH-----1193
QY 1607 NAKPALDINWPSRPRTFGITWRYLQTVKSL--AGVSLMLRLWLASLRWDDMAAKVPPG 1664
DB 1194 ---SLDKTSHS---TGLAFLSLSAPSSSSSPGAGVSSYIS-----QPG 1231
QY 1665 GGSTRTETSETEITTEIIKRRDVGPIRPEYCIIRKIICPIGVPEPKETP--TPQRKGL 1723
DB 1232 G-----LHPLVIPSVIASTPILPQAAGP 1254
QY 1724 RSSALRPKRPTPKOTGPV-----IETWVAEE-----ELE-----1754
DB 1255 TSTPLLPQVSIPLVQPVANVPVQOQLIHSQOPALLPNQPHTHCEVSDSDTOPKAPG 1314
QY 1755 LWEIRAFARVE---KEKAQAVEQAKKRLBQO---KPTV---IATSTTSTP---SSPTS 1802
DB 1315 IDDKTLEELKRSLEFSEHSSGAQVSVLSLSTVIESTVTPGPTTAVARSKLLTSITS 1374
QY 1803 TISPAQKV---MVAPISGVTTGKMYLVLTTKVGSFATVTFQONKNFHQTATVWVKQQS 1858
DB 1375 TCLPPTNLPLGTVALPVPVVTGP-----QVSTPVSTT-----1407
QY 1859 NSGVVQVQOKVLGIIPSTGCTSQOFTSFQPTATVTIRPNTSGSGGTTNSOVITGPQI 1918
DB 1408 -----TSGVKPGTAPSKPPLTKAPVLPGVTGELPAGLPLEQPLPFPF--1448
QY 1919 RGMVTIVRTPLOO-----STLGRKAIIRTPVMVQPGA--PQVMTQIIR 1959
DB 1449 GPSLTQSOQPLEDLDAQLRRTLSPEXITVTSAVGPSMAAPTATTEAGTQPKQGVQVKE 1508
QY 1960 QOPVSTAVSA-----PNTVSTTPGQKS-----LTSATSTNTOSSASQPP 1999
DB 1509 GPVLATSSGAGVFKMGRFQVSVAAQGAQKEGKNKSEDAKSVHFSESTSESSVLSSSS---1565
```


QY 622 GEFKNGSVHGSKVLITSTLRLTITOLENNIPSPSELHPNASHANWIKAV----- 672
Db 362 GSVIGD-----LGLATK-----RASFAKSVIGTFEPMAP 392
QY 673 QMCKSPRE-----PALALALECA-----VKPV-----VMLPIW 701
Db 393 EMYEKYDESVDVAFGCMLEMATSEYPYSECQNAQIYRRVTSYGVKPASFVKVAPEV 452
QY 702 RE-----FLGTHRLHMTSI-----EREKEK-----VKKK 727
Db 453 KEIIEGCIQONKDERYSIKDLNLNHAFOEETGVRVELAEEDGKIAIKLWLRIEDIKKL 512
QY 728 EKKOEETMQOATWVKYTFPVKHQVWKQKEEYRVGY-----GWSWLSKTHVTRFVKPL 784
Db 513 KGKYKDNFA-----IEFSDLEKRDVPEDVAGVMEVSGYCEGDHKTAKA----- 557
QY 785 PGNTNVNRYKSLGKTKNMNDENMDESOKRKSRSPPKKIKIBPDSEKDEVKGSDAAGADQ 844
Db 558 -----IKDRVSLIKRKRQOLVREEQEK-----KQESSLKQOQVESSASQ----- 600
QY 845 NEMDISKITEKKDQDVRELLSDSDKCKEPMEDVDDMKTESHVNCQESSQVDVNVNSE 904
Db 601 -----TGIIKOLPASSTGIP-----TASTTSASVSTQVPEEPEA 634
QY 905 GFHLRTSYKKTCKSKLDGLLERIKQFTLEEKQORLEKIKLEGGIKGIGKTSNSKNLS 964
Db 635 DOHQLOLQOQPSISVLSDGTVD-----SGQSSVFTESRVSS 671
QY 965 ESPVITRAKEGCQSD-----SMRQEQSPNANDQPEDLLQGCQSQDSSVLRM 1011
Db 672 QQTV-----SYGQHEQAHSHTGVPCHIPSTVQAQSQPHGVYPPSSVAAQSQGQSPSSSL 727
QY 1012 SDPHTTNKLYPKORVLDVDSIRSPETKCPKQNSIENDI-----EKVSDLASRGOEP 1064
Db 728 TGVSSSOPIQHPOQO-----QGIIQQTAPPQQTQVQYSLQTSSTSEATTQAQVSPQOAP 780
QY 1065 T-----KSKTKGNDOFFIDDSKLASADDITGLICKNKKPLIQEESDTIVSSSKSALHS 1116
Db 781 QVLQVQVAGKOSTQG-----VSQVAPAEVAV-----AQD-----QATQPTLLASSVDSAH 827
QY 1117 SVPKSTNDRTAPLSRAMDFEGK-----LGCDSSESNTL 1150
Db 828 DVASGMSDGNENVPSSGRHSGRTKRRHYKSVRSRSHRHKTSRPLRLILNVSNKGRDV 887
QY 1151 ENSSDT-----VSTQ-----DSSEE-----DMIVONSNESTSEFRTREQDVEVLEP 1192
Db 888 ECQLETHNRKMVTFKFDLDGDNPEEATIMVNDFFILAIERESFVDQVR-----EIIERKADE 944
QY 1193 LKCELVSGESTGNCEDRLPVKGTANGKKPSQOKKLEER-----PVNKCSDQIKLANWTDDK 1249
Db 945 MLSEDDVSVEPEGD-----QGLSEUQKODYGFSGSKLEGEFKQIPASSMPQQIGIPT----- 998
QY 1250 NENRESEKKGORTSTFQINGKDNKPKIYKAGE-----CLKEISESRVSGNVPEKVNINKI 1307
Db 999 -----SSLTQV-----VHSAGRFRIVSPDESRLRESKVEPP----- 1029
QY 1308 IPENDIKSLVKGESAIRPFINGVIMEDFNERNSETKSHLLSSDAEG-----NVRDS 1361
Db 1030 ---SEITDTVAASTAQSPGMN-----LSHSASSLSLQQAFLSELRAQMTGPNATPNFSHT 1083
QY 1362 LETLPSYK---ESDSTQTTTPSACSPESNSVNOVEDMEIETSEKVKYTSSTPITSEESNL 1418
Db 1084 GPTFFVPPFLUSSITAGVPTAAATAPVATSSPPND---ISTS-----VIQSEVTVPTBEGI 1137
QY 1419 -----SNDFIDENGLIPKNKNENVESKRRKTVITEVTTMTSTVATESKTVIKVEKGDKQT 1473
Db 1138 AGVATSGVTSGLPI-----PPVSESPVLSVSVSSIIPAVVSVISTTSPSLQVPTSTSEI 1194
QY 1474 VVSSTENCAKSTVTTTTTTTTVKLSTPSTGGSDVSIISVKEQSKTVTTTTVTVDLSLTGTGTL 1533
Db 1195 VVSSTALYPSVTVSATS-----SAGGSTATPGPKPPA---VVSQQAAGS---TTVGATL 1243
QY 1534 VTSMTVSKEYSTRDKVXKLMKFSRPKKTRSGTALPSYRKFTVTKTKKSIFVLPNDDLKKLA 1593

Db 1244 TSVSTTTSPFPASOLS-----QUSSTSTPTLAETVVVSAH----- 1281
QY 1594 RKGIREVPYENYNAKPALDIWPYSPRPRTFGITWRYRLQTVKSL---AGVSLMLRLLLWAS 1651
Db 1282 -----SLDKTSHSS---TTGLAFSLSPSSSSSPGAGVSSYIS----- 1316
QY 1652 LRWDDMAKVPDPGGSTRTTSETTEITTEIKRRDVGYPGIRFEYCIRKIKICIPGVPET 1711
Db 1317 -----QPGG-----LHPLVIPS 1329
QY 1712 PKETP-TPORKGLBSSALRPKRPETPKOTGPV-----LIETWAAE----- 1751
Db 1330 IASTPILPQAAAGPTSTPLLPVPSIPPLVQPVANVPVAVQOFLIHSQOPALLPNOPHTHC 1389
QY 1752 -----LWEIRAFARVE---KEKAQAVEQOAKRLEQO---KPTV---IATS 1792
Db 1390 PEVSDTOPKAPAGIDDDIKTLEKLSLRFSEHSSGCAHASVLETSLVIESTVTPGIPPT 1449
QY 1793 TTSPT---SSTTSTISPAKV---MVAPISGVTGTKMVLTKKVGSPATVTTQOQKNF 1845
Db 1450 AVAPSKLLTSTSTCLPPTNLPLGTVALPVTVPVTPG-----QVSTPVSTT----- 1495
QY 1846 HQTATWVKOGSNGSVVVOOVKVLGIIPSTSTGTSQOTFTSFQRPETATVTRPNTSGSG 1905
Db 1496 -----TSGVKPGTAPSKPPLTKAPVLPGVTELPAG 1525
QY 1906 TTSNSQVITGPQIRPGMTVIRTPLOQ-----STLGKAIIRTPVMVQPG 1948
Db 1526 TLPSEQLPPFP--GPSLTQSOQPLEDLDAQLRRTLSPKXITVTSAVGPVSMAPTAITEA 1583
QY 1949 A--POQVMTQIIRQPVSTAVSA-----PNVVSSTPGQKS-----LTSAT 1986
Db 1584 GTQPKGVSVQKEGVPVLATSSGAGVFKMRFQVSVAAQGAQKGNKSEKADAKSVHFEST 1643
QY 1987 STSNIOSSASQPPRPOQGVKLTMAQLTQLTQGHGNOGLTVIOGQGTTLQQLIPIQG 2046
Db 1644 SESSVLSSSS-----PESTLVKPE-----PNG 1665
QY 2047 VTVLPGGQOOLQMAAMPNGTVQREFLTPLATTATTASTTT-----TTVSTTAAGTGE 2098
Db 1666 ITI-PGISSDVPESAH-----KTTASEAKSDTGQPTKVGREFQVTTTANKVGR 1711
QY 2099 QRSK-----LSPQMVHODKTLPPAQSSSVGPAKAQPOTAQPSARPQQT 2144
Db 1712 FSVSKTEDKITDTKKEGVASPPFMDLEQAVLP-----AVIPKKEKPELSEPHLNGPSS 1766
QY 2145 QPO-----SPAQP----- 2152
Db 1767 DPEAFLSRVDVDDGGSPHSPHQSLSSKSLPSONLSQSLNSFNSSNMSDNESDIEDDL 1826
QY 2153 -----EVQ-----TQPEVQTQTTVSSHVP-----SEAQPTHAQS 2181
Db 1827 KLELRLRDKHLKEIQDLQSRQKHEIESLYTKLGKVPVPAVIPPAAAPLSGRRRRPTKSKG 1886
QY 2182 SKPOVAA---OSQPO---SNVQOQSPVRVQSPQSRIRPSTPSQLSPGQSQVQTTTSQPI 2236
Db 1887 SKSRSSSLGNKSPQLSGNLSCQSAASVHLHPQOT-LHP-----PG----- 1925
QY 2237 PIQHTSLQIPSQGPO-SOPQVOS-SQTQLSSGOTLN-OVSVSSPSRQLQIOQP 2289
Db 1926 -----NIPESQNLQLLPKPSPPSNDNIYSAFTSDGALSVPFSLSPAGOGIKOP 1973

RESULT 10
US-09-854-856-62
; Sequence 62, Application US/09854856
; Patent No. 65411252
; GENERAL INFORMATION:
; APPLICANT: Walker, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.

Qy	1117	SVPKSTNDRDATPLSRAMDFEGK-----LGCDSSESNSTL	1150
Db	768	DVASGMSDGNENVPSSGRHEGTRKRRHYKSVRSRHEKTSRPLRLINLSNKGDRVV	827
Qy	1151	ENSSDT-----VSIQ-----DSSEE-----DMIVQNSNESISBQFRTREQDVELEPP	1192
Db	828	ECQLETHNRKRWTFKFDLGDNDPNEETATIMVNNDFILATERESFVDQVR-----EIEKADE	884
Qy	1193	LKCELVSGETGNCERDLRPLVTEANGKKPSQOKKLEER---PVNKCSDQIKLKNYTDKK	1249
Db	885	MLSEDSVSEPEGD-QGLESIQGDDYGFSGSQLEGEFKQIPASSMPOQIGIPT-----	938
Qy	1250	NENRESEKKQRTSTFQINGKDNKPKIYLKGE--CLKEISESRVSGVGNPEKVNINKI	1307
Db	939	-----SSITQV-----VHSAGRFFIVSPVESRLRESKVFP-----	969
Qy	1308	IPENDIKSLTVKESAIRPFINGDIVIMEDFNERNSSETKSHLSSSDAEG-----NYRDS	1361
Db	970	---SEITDTVAASTAQSPGMN---LSHSASSLSLQAFSELRRRAQMTGEPNTAPPNFSHT	1023
Qy	1362	LETLPSTK---ESDSTQTTTPSASCPSNSVQVEDMEIETSEVKKVTSPITSEESNL	1418
Db	1024	GPTFPVVPFLSLGIAGVPTTAAATAPVPATSSPND--ISTS-----VIOSEVTVPTTEGI	1077
Qy	1419	-----SNDFTDENGLPINKNENYNGSRKRTVITEVTTMTSTVATESKTVIKVEKGDQOT	1473
Db	1078	AGVATSGVTVSGGLPI---PPVSESPVLSSVVSSITIPAVVISITSPSLQVPTSTSEI	1134
Qy	1474	VVSTENCASKTVTTTTTKLSTPSTGSGVDIISVKEOSKTVVTTVTDSITTTGGTL	1533
Db	1135	VVSTALYPSVTSATSA-----SAGGSTATPGPKPPA--VVSQQAAGS-TTVGATL	1183
Qy	1534	VTSTVTSKEYSTRDKVLMKFSRPKKTRSGTALPSYRKFTYKTKKSIFVLPNDLKKLA	1593
Db	1184	TSVSTTTSFSTASQLSI-----QLSSSTSTPTLAETVVVSAH-----	1221
Qy	1594	RKGGIREVPVFNYNNAKPALDIWPPSPRPTFGITWRYRLQTVKSL--AGVSLMLRLLLAS	1651
Db	1222	-----SLDKTSHS--TTGLAFSLSAPSSSSSPGAGVSSVIS-----	1256
Qy	1652	LRWDDMAKVPVPGGGSTRTETSETTEITTEIIRKRDVGPYIGIRFECYIRKILCIGVPET	1711
Db	1257	-----QPGG-----LHPLVTPSV	1269
Qy	1712	PKETP-TPQRKGLRSSALRKRPRPKOTGPV-----IETWVAEE-----	1751
Db	1270	IASTPILPQAAGTSTPELLQOVPSIPPLVQPVANVPAVQOTLIHSQOPALLNQPHTHC	1329
Qy	1752	-ELE-----LWEIRAFARVE---KEKAQAVEQQAQRLEQQ---KPTV---IATS	1792
Db	1330	PEVSDTQPKAPGIDDIKTLEKRLSLFSEHSSGGAHASVLSLETSLVIBESTVTPGIPTT	1389
Qy	1793	TTSTP---SSTTSTISPAKV---WVAPISGVSITTKMVLTKVCSPATVTFQONKNF	1845
Db	1390	AVAPSKLLTSTSTFCLPPTNLPLGTVALPVTPVTPG-----QVSTPVSTT-----	1435
Qy	1846	HOTFATVVKQGSNSGVVQVQOKVLGIIPSTGTSQOTFTSFOPRTATVTRPNTSGSG	1905
Db	1436	-----TSGVKPGTAPSKPLTKAPVLVPGTELPA	1465
Qy	1906	TTSNSQVITGPQIRPGMTVINTPQQ-----STLKAIRTPVMVQPG	1948
Db	1466	TLPEQLPPPF--GPSLTQSQPLEDLDAQLRRTLSPXITVTSVAVGPVSMAPATAITEA	1523
Qy	1949	A--PQOVWTOIIRQIPYSTAVSA-----PNTVSTTPGOKS-----LTSAT	1986
Db	1524	GTQPKGVQSVKEGVLVLTSSGAGFVMGRFRQVSVAADGAQKGEKKNKSEDAKSVHFESST	1583
Qy	1987	STNSIQSASOPPRPOQGVKLTMAQLTQTHGGNGGLTVVITQOGQTTGQLQLIPQG	2046
Db	1584	SESVLSSSS---PESTLYKPE-----PNG	1605

Qy	2047	VTVLPGPGQOLMQAAMPNGTVORFLFTPLATTATATATTT-----TTVSTTAAGTGE	2098
Db	1606	ITI-PGISSDVPESAH-----KTTASEAKSDTGQPTKVGRFQVTTTANKVGR	1651
Qy	2099	QRQSK-----LSPQMVOHQDKTLPPRAOSSVGPAAKQPOQAQPSARQPOT	2144
Db	1652	FVSKTETDKITDKKEGPVASFPMDEQAVLP-----AVIPKKEPSELSEFSLHNGPSS	1706
Qy	2145	QPO-----SPAQP-----	2152
Db	1707	DPEAAFLSRDVGDSGSPHSPHOLSSKSLPSQNLSSLSNFSNSSYMSDNESDIEDL	1766
Qy	2153	-----EVO-----TQPEVQOTQTVSSHVP-----SEAQPHAOS	2181
Db	1767	KLELRRLRDKHLKEIQDLOSROKHETESLYTKLGKVPVAVIIPPAAPLSGRRRRPTKSG	1826
Qy	2182	SKPQVAA---QSQPO---SNVOGOSPVYVQSPSTRIRPSTPSQLSPGQSQVQVTTTSQPI	2236
Db	1827	SKSSRSSLGKNSPQLSGNLSGQSAASVLPQOT-LHP-----PG-----	1865
Qy	2237	PIQPHTSLOIPSGOQPQ-SQPQVS-STQTLSSGQTLN-QVSVSSPSRPLQIQOP	2289
Db	1866	-----NIPESQONLLOPLKPSPSNDNLYSAFTSDGAISVPSLSAPQOGIKQP	1913

RESULT 9

US-09-854-856-16

Sequence 16, Application US/09854856

Patent No. 6541252

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Hilbun, Erin

APPLICANT: Donoho, Gregory

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides

TITLE OF INVENTION: Encoding the Same

FILE REFERENCE: Lex-0178-USA

CURRENT APPLICATION NUMBER: US/09/854,856

PRIOR FILING DATE: 2001-05-14

PRIOR APPLICATION NUMBER: US 60/206,015

PRIOR FILING DATE: 2000-05-19

NUMBER OF SEQ ID NOS: 64

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 16

LENGTH: 1999

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)..(1999)

OTHER INFORMATION: Xaa = Any Amino Acid

US-09-854-856-16

Query Match 2.5% Score 358.5; DB 4; Length 1999;

Best Local Similarity 18.2% Pred. No. 9.8e-13;

Matches 415; Conservative 292; Mismatches 722; Indels 847; Gaps 95;

399 TNKARGSNKSFALAAANEIILSIRAKKGDIDNVKSPPEETEKDKNENETENDSKDAEKNEEF

450

160 TSKDRPVQPSLVGSKKEPPPPARSGGG-----SAKEPQERSQQQDDI-----EEL

207

459 EDOSLEKSDDKTPDDDPQEGKSEVDFKSEKNGELSESPGAGKAGSGSTRI-ITRLRN

517

208 ETKAVGMSNDGRFLKFDIEIGR---GSFKT-----VYKGLDTEVTEVAVWCEL

252

518 PDKLSOLKSOQVAAAAAHEANKLFFKEGKEVLVWNSQGEISRLSTKKEVIMKGN--INNYF

575

253 QDRKLITSEQR-----FKEEAEMLKGQLQHNIVFRYDVSWEVSTVKGKCVLVT

301

576 KLGQEGKYRYVHYNQYTSNFSALNKHQHRED-----HDKRRHLAHK-----FCLTPA

621

302 ELMTSGTLTKYLRKFKVMKIKVLRSWCRQILKGLQFLHTRTPPIIHRDLKCDNIFITGPT

361

Db 5323 SMQTLRLVNNQNAIHSYFNEDSTQKNTYDNDNGSTYITGQHNPELNKSTIDOTI 5382
QY 2111 HQDKYLPQAQSSVGPAAQQTQAPSPQPOQPOQSPAPQEVOTQPEVOTQTV--SS 2168
Db 5383 SRINT--AKNDLHGVEKLQDKG-----TANQIEIGOLGYNLDPQKSGEESLVNGS 5430
QY 2169 HVPSEAPTHAQSCKPQVAQSQPQSNVQSGSPRVQSPQSPQTRIRPSTPQSPQSQSQV 2228
Db 5431 NTRSEVE-EHLEAKSLNNAMKQURDKVAETKNYK---QSSDYINDSTEHORGVDQALQE 5486
QY 2229 QTTTSOPIPIQHTSLQIPSGQPQ-SOPQVQSSTQTLSQQT----- 2270
Db 5487 AENIINEI-----GNFTLNKSEIEQKLQQLTDAQNALQSHLLEAKNNAIT 5533
QY 2271 -LNOVSVSSPSRPOLOIQOPQPVIAVPQLOQQVQVLSQIQS---QVVAQIAQAGSGVPQ 2327
Db 5534 GINKLTALNDAQROKAIENVAQAOQ-TIPAVNQOQLTLREINTAMQALRDVKVGOQNNVHQ 5592
QY 2328 I-----KLOLPQIQQSSAVQTHQIQNVVTVQAAASVQEO 2362
Db 5593 SNYFNEDEQPKHNTYDQVQAGOTIIDLQDDPMKNKEIQAININTQT--ALUSGENKL 5650
QY 2363 QRVOQLRDQO---QKKKQOQIEIKREHTLQASNOSEIIQKVVMKH-NAVIEHLKO--- 2414
Db 5651 HTDQESTNRQIEGLSSLNTAQINAEKDLVNOAKTRTDVQAKLAAAKAINSAMSRLDGIQ 5710
QY 2415 -----KKS-----MTPAEREENQRMIVCQVMKYILDKIDKEEQAARKRRE-- 2457
Db 5711 NKEDIKSSAYINADPTKVAYDQALQNAENIINATPNVELNKATIEQALSrvQQAQOQL 5770
QY 2458 ESVEQ-KRSKONATK---LSALLFKHKEQLRAEILKRALLDKLDQIEVQEEELKRDKI 2512
Db 5771 DGVQOLANAKQOATQTVNGLSNLDGQKRELNL-----LINSANTRTKVQEEELNKATEL 5824
QY 2513 KKEKDLM-----QLAQATAVAAPCPP-----VTPVLP----- 2539
Db 5825 NHAMEALRNSVQNDVQVQSSNYVNEDEQPEQHNTYDNAVNEAQATINNAQPVLDKLAIER 5884
QY 2540 -----APPAPPPPPPGVQHTGLLSTPTLPVSAQKRRKEEDSSSKS 2595
Db 5885 LTQTVNTTKDALHGAQKLTQDQAAETGIR--GLTS-----LNEPQKNAEVAKVTAATT 5936
QY 2586 KKKMISTSKETKDKTLYCICKTPYDE-----SKEYIGDRCONVTHGRCVGLQSEAE 2640
Db 5937 RDEVNTRQEAITLDTAMLGRKSIKDKNTKNSKYINEDHQOQOQAYDNVNNAQ---Q 5993
QY 2641 LIDYVCPQOQSTEDAMTVTLPLT-----EKDYEGLRVYLSLQAHKMAWPF 2687
Db 5994 VIDE---TQATLSSDTINQLANAVTQAKSNLHGDTKLQHKDSAKQITQALONLSAQKH 6050
QY 2688 LEPVDPNDAPDYGVIVKEPMDLATMEERVQRRYIE-----KLTEFVADMTKIFDNC 2738
Db 6051 ME-----DSLIDNESTRTQVQHDLTEAQALDGLMGALRESIKDYNIVNSG 6096
QY 2739 RYVN-PSDSPFYQCA 2752
Db 6097 NYINAEPSKKQAYDAA 6112

RESULT 8
US-09-854-856-48
; Sequence 48, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Waikie, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: Encoding the Same
; CURRENT APPLICATION NUMBER: LEX-0178-USA
; CURRENT FILING DATE: 2001-05-14

; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1939)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-48

Query Match 2.5%; Score 358.5; DB 4; Length 1939;
Best Local Similarity 18.2%; Pred. No. 9.3e-13;
Matches 415; Conservative 292; Mismatches 72; Indels 847; Gaps 95;

QY 399 TNKARGSNKSFLLAANEELIESIRAKKGDIDNVKSPETEEDKDKNETENDSKDAENREEF 458
Db 100 TSKDRPVSPSLVSGSKKEPPPARSGGG-----SAKEPQERSQQQDDI-----EEL 147
QY 459 EDQSLEKSDSDKTPDDDPQEGKSEVGFQKSEKSNGLSESPCAGKAGSGSTRI-ITRLRN 517
Db 148 ETKAVGMSNDGRFLKFDIEIGR---GSFKT-----VYKGLDTETTVAVMCEL 192
QY 518 PDSKLSOLKSOQVAAAAHEANKLFKEKEVLVNVQSGEISRLSTKKEVIMKGN--INNIF 575
Db 193 QDRKLTKSERQ-----EKEBAEMLKGLQHPNVIRFYDSWESTVKGKCIVLVT 241
QY 576 KLQEGEKYRVYHNOYNTNSFALNKHQHRED-----HDKRRHLAHK-----FCLTPA 621
Db 242 ELMTSGTLKTKYLRKFKYMKIKVLRSMWCRLKGLQFLHTRPTPIIHRDLKDCDNIFITGPT 301
QY 622 GEFKWSGVHSGKVLITLITLITLITLITLITLITLITLITLITLITLITLITLITLITLIT 672
Db 302 GSVKIGD-----LGLATLK-----RASFAKSVIGTPEPMAP 332
QY 673 QMCKPRE-----FALALAILBCA-----VKPV-----VMLPIW 701
Db 333 EMYEEKYDESVDVYAFGCMCMLEMATSEYPYSECNAQIYRRVTSVGVKPFASFDKVAIPEV 392
QY 702 RE-----FLGHTLHRWTSI-----EREKEK-----VKKK 727
Db 393 KEITEGCIQNKDERYSIKDLLNHAFTQEEFGVRAVEAEDDEGKIAIKMLRIEDIKKL 452
QY 728 EKKQEEETMQAATWVKYTPPVKQVWVKQKGEERYVTGY---GQWSWISKTHVYRFVPKL 784
Db 453 KGKYKDNEA-----IEFSFDERDPEDVQAQEMVESGYVCEGDHKTWAKA----- 497
QY 785 PGNTNVNVRKSLGKTKNNMDENMDSEKRCSPKKIKIEPDEKSEDEVKGSAAKQADQ 844
Db 498 -----IKDRVSLIKRKRQRLVREEQKK-----KQESSLKQOVQESSASQ----- 540
QY 845 NEMDISKITEKKQDQVKELDSDSDKPCKEEPNEVDDMDKTESHVNCQESSQDVVVNSE 904
Db 541 -----TGIKQLPSASTGIP-----TASTTSASVSTQVPEPEPEA 574
QY 905 GFHLRTSYKKTKSKSLDGLLERRIKOFTLEEKORLEKIKLEGIGIKGTSTSNKSLNS 964
Db 575 DQHQOLOYQOPSISVLSGDIVD-----SQGSSVFTESRVSS 611
QY 965 ESPVITKAKGCGQSD-----SMRQOSPNANNDQPEDLQGGCSQSSSVLRM 1011
Db 612 QQTIV-----SYGSHQEAHSTGTVPGHIPSTVQAQSQPHGVYPPVPSVAQSQSQSSSSL 667
QY 1012 SDPSHTTNKLYPKDRDLDDVIRSPTKCPKQNSIENDI-----BEKVSDLASRQOEP 1064
Db 668 TGVSSSOPIQHPQOQ-----QGIQQTAPPQQTQVYLSQTSSTSEATTAPQVSPQAP 720
QY 1065 T-----KSKTKGNDFFIDDSKLASADDITGLICKNKKPLIQEESDTIVVSSSKSLHS 1116
Db 721 QVLPQVSAGKQSTOG-----VSQVAPAEPVAV-----AQP--QATQPTTLASSVDSAH 767


```
Db      828 DVASGMDGHNVPSSGRHEGRTTKRHYKSVRSRSHKTSRPKRLILNVSNKGRV 887
Qy      1151 ENSSDT-----VSIQ-----DSBE-----DMIVQNESISEQFRTREQDVEVLEP 1192
Db      888 ECOLETHNRKMWTFKFDLGDNDPEIATIMVNDPILAIERESVDQVR--EIIKRADE 944
Qy      1193 LKCLVSGESTGNCEDRLPKVGTANGKPKSQKKLEER---PVNKSDQIKLNTDKK 1249
Db      945 MLSEDSVPEPGD--QGLLESQGGDDYGFSGOKLEGEFKQIPASSMPOQIGIPT----- 998
Qy      1250 NENRESEKKGORTSTFOINGKDNKPKIYLKGE--CLKEISESVRSVGNVEPKVNNINKI 1307
Db      999 -----SSLQV-----VHSAGRFRIVSPESKRESKVP----- 1029
Qy      1308 IPENDIKSLAVKESAIRPFINGDVIMEDFNERNSETKSHLLSSSDAEG-----NYRDS 1361
Db      1030 ---SEITDVAASTAQSPGMN---LSHSASSLSLQQAFFSELRAQMTGPNAPPNFSHT 1083
Qy      1362 LETLPSTK---ESDSTOTTPSACPSNSVQVEDMEIETSEVKKVTSSPITSEESNL 1418
Db      1084 GTPFPVVPFLSLSIAGVPTTAATAAPATVPATSSPND--ISTS-----VIQSEVTVPTTEGI 1137
Qy      1419 -----SNDFIDENGLPKNNVNGESKRKTIVITEVTTMTSTVATESKTIVIKVEKDKQT 1473
Db      1138 AGVATSGVTVSGGLPI---PPVSESPLVSLSVSSIIIPAVVISISTSPSLQVPTSTSEI 1194
Qy      1474 VVSSTENCAKSTVTTTIVTKLSTPSTGSGVDIIISVKEQSKTVTTTIVTDSLTSTTGGTL 1533
Db      1195 VVSSTALYPSVTVSATS-----SAGGSTATPGPKPPA--VVSQQAAGS--TTVGATL 1243
Qy      1534 VTSMTVSEKSTRDKVKLMKFSRPRKTRSGTALPSYKFKTKYKTKKIFLVPNDLKKLA 1593
Db      1244 TSVSTTTFSPSTASQLSI-----QLSSSTSTPLAETVWSAH----- 1281
Qy      1594 RKGIREVPVFNYNAKPALDWPVSPRPTFGIWRVRLQTVKSL--AGVSLMLRLWLAS 1651
Db      1282 -----SLDKTSHS---TTGLAFSLASPSSSSPGAGVSSYIS----- 1316
Qy      1652 LRWDDMAKVPVPGGGSTRTETSEITEITTEIKRDRVGPYRFEYCIKRIICPIGPVET 1711
Db      1317 -----QPGG-----LHPLVIPS 1329
Qy      1712 PKETP--TPQRKGLRSSALRKRPTPKQTGPV-----IETWVAEE----- 1751
Db      1330 IASTPILPQAAGPTSTPLLPQVPSIPPLVQPVANVPVAVQOTLIHSQPALLNPQPTH 1389
Qy      1752 -ELE-----LWEIRAFARVE---KEKAQAVEQAKRLEQ---KPTV--TATS 1792
Db      1390 PEVDSOTQPKAPGIDDKTLEELKRLSLFSEHSSGAQHASVLSLTSVISTVTPGIPPT 1449
Qy      1793 TTSPT-----SSTTSPSAQV---MVAPISGVTTGTGMVLTKVGSPTATVTFQONKNF 1845
Db      1450 AVAPSKLLTSTSCLEPTPLGLTVALPVPVTPG-----QVSTPVST----- 1495
Qy      1846 HQTFATVWKQOSNGVQVQKVLGIIPSTGTSQOTFTSFQPRATVIRPNTSSGG 1905
Db      1496 -----TSQVKGPGTAPSKPPLTKAPLVPVGTLPAG 1525
Qy      1906 TTSNSQVITGPQIRPGMTVIRTPLQ-----STLGKAIIRPVMVOPG 1948
Db      1526 TLPSEQLPPP--GPSITQSQOPLDLDQILRRLTSLPEXITVTVSAVGFVMAAPTATEA 1583
Qy      1949 A--PQVMTQIIRGQPVSTAVSA-----PNTVSTPQOKS-----LTSAT 1986
Db      1584 GTQPKGVQVQKGVPLATSSGAGVFKMGRFOVSAADGAQKEGKNKSEDAKSVHFSST 1643
Qy      1987 STSNIQSSASQPPRQGVQKVLWALQTLTQGHGNGQLTVVIQGGQTTGQLLIPQG 2046
Db      1644 SESSVLSSSS-----PESTLVKPE-----PNG 1665
Qy      2047 VTVLPGGQQLMOAMNGTVQRFLETPPLATTATTAFTTT-----TTVSTTAAGTGE 2098
```

```
Db      1666 ITI-PGISSDVPESA-----KTTASEAKSDTQPTKVGRFQVTTTANKVGR 1711
Qy      2099 QROSK-----LSPQMVDHDKTLPPQASSSVGPAKAQAPQTAQPSARPOQT 2144
Db      1712 FSVSKTEDKITDKKEGVPASPPMDLEQAVLP-----AVIPKKEKPELSEPHLNGPSS 1766
Qy      2145 QPQ-----SPAQP----- 2152
Db      1767 DPEAAFLSRDVGSGSPHQLSSKSLPSQLSLSNSFSNSYSMSDNESDIEDL 1826
Qy      2153 -----EVQ-----TOPEVOTQTVSSHP-----SEAOPTHAQS 2181
Db      1827 KLELRRLDKLHEIQDLQSRQKHEIESLYTKLGKVPVAVIIPPAAPLSGRRRPTKSG 1886
Qy      2182 SKQVAA---OSQPQ---SNVQGSQPVRSQPSQTRIRPSTPSQLSPQQSQVQVTTSQPI 2236
Db      1887 SKSRSSSLGNKSPQLSGLNSGQSAASVLPQQT-LHP--PGNIPESGQNL-----LQPL 1939
Qy      2237 PIQPH-----SLOIPSGQGPQSQPVQSSTQTLSSGQTLNQVSVSSPSRPLQ 2285
Db      1940 KPSPPSDNLYSAFTSDGAISVPSLSAP---GGQTSSTNTV--GATVNS-----Q 1983
Qy      2286 IQQPOP 2291
Db      1984 AQAQP 1989

RESULT 7
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match      2.5%; Score 363.5; DB 4; Length 10182;
Best Local Similarity 17.3%; Pred. No. 5.2e-12;
Matches 573; Conservative 488; Mismatches 1236; Indels 1019; Gaps 142;

Qy      136 VSOEECTLMAEMHVLLKAVLREDDTSNTTFGPADKDSVNSTLYFDGMTWPEVLR-VY 194
Db      3117 IENQYNTAKNEAHNVL-----ENTNPTVNAVEDALRKINA-----IQEVTTRAIN 3161
Qy      195 CESDKYHHVLPYQEAEDYPYGVENKIKVLQFLVDQFLTNTIAREELMSEGVQYDDHC 254
Db      3162 ILQDKEDNSL-----VRAKEK-----LDQAINSQPSLNGMTQESINNYTKR 3204
Qy      255 RVCHKGLDGLCCETCSAVYHLECVKPLPEVPEDEWQCEVCV-----AHKVPGVTVDCV 307
Db      3205 REAQNTAS-----SADTIINNGDASIEQITENKIRVEEATNALNEAKQHLTADTTSLK 3257
Qy      308 AETQKNKPIRHEPIGVDRSRKRYWFLNRLRIIEEDTENENKIKWYTSKVQLAEILDC 367
Db      3258 TEVRK-----LSRR-----GDTNNKKPSSVSAYN-----NTIHS 3286
Qy      368 LDKDYWEAE-----LCKILEMREEHHRMDITEDLTN----- 400
Db      3287 LQSEITQENRANTINKPIRSVEEVNNAHEVNQLNQRUTDITINLLOPLANKESLKEAR 3346
```

QY 1712 PKETP--TPQKGLRSSALRPKRPKOTGPV-----IETWVAEE----- 1751
Db 1270 IASTPILQAAAGTSTPLLPQVPSIPPLVQPVANPAVOOTLIHSQOPALLNPQPHTHC 1329
QY 1752 -ELE-----LWEIRAFARVE---KEKAQAVEQAQKLEQO---KPTV---IATS 1792
Db 1330 PEVDSOTOPKAPGIDDIKTLEKRLSLFSEHSSGAHASVSLETSLVIESTVTPGIPPT 1389
QY 1793 TTSPT---SSTSTISPAKV---MVAPISGVTTGKWLTKVGVSPATVTFQONKF 1845
Db 1390 AVAPSKLSTSTCLPTNLPLGVALPVPVTPG-----QVSTPVST----- 1435
QY 1846 HOTEATWVKQSGSVQVQOQKVLGIIPSTSTGTSQOTFTSPQRTATVIRPNTSGSG 1905
Db 1436 -----TSGVRPGTAPSKPPLTKAPVLPVGTLPAG 1465
QY 1906 TTSNQVITGPQIRGCMVIRTPLOQ-----SILGKAILIPTVMVQPG 1948
Db 1466 TLPSEQLPFP--GPSLTQSOQPLEDLDAQLRRTLSPEXITVTSVAGPVMAAPTATEA 1523
QY 1949 A--POQVMTQIIRGQPVSTAVSA-----PNTVSTPGOKS-----LTSAT 1986
Db 1524 GTQPKGVSVQKGVPLATISSAGVFKMGRFOVSVAAQNAQKGNKSEDAKSVHFEST 1583
QY 1987 STSNTOSSASQPPRQOQGVKLTMAQLTQLTQGHGNGOGLTVVIOGQGTQLOLIPQ 2046
Db 1584 SESSVLSSES-----PESTLVKPE-----PNG 1605
QY 2047 VIVLPGQOQLMAAMPNGTVQRFLETPLATTTASTTT-----TTVSTTAAGTCE 2098
Db 1606 ITI--PGISSDVPESAH-----KTTASEAKSDTGQPTKVGRFQVTTANKVGR 1651
QY 2099 QROSK-----LSPQOVHODKTLPPAQSSSVGPAKAQOPOQAQSRPQOT 2144
Db 1652 FSVSKTEDKITDKKEGVPVSPFMDLEQAVLP---AVIPKKEPSEPSHLNGPSS 1706
QY 2145 QPO-----SPAQP----- 2152
Db 1707 DPEAPFLSRDVGSGSPHSQLSKSLPSONLSQSNFNSVMSSDNESDIEDL 1766
QY 2153 -----EQV-----TOPEVOTQTVSSHPV-----SEAQPHAS 2181
Db 1767 KLELRRLDKHLKEIQDLSRQKHETESLYTKLGVPPAVIIPPAAPLGRRRRPTKSG 1826
QY 2182 SKPQVAA---SQSQ---SNVQSQSPVRVQSPQTRIPSTPSQLSGQSQVQTTSQPI 2236
Db 1827 SKSRSSSLGNKSPQLSGNLGSAASVLPQQT-LHP--PGNIPESGQNL---LQPL 1879
QY 2237 PIQPH-----SLQIPSOQOPQSQPOVQSQTQLSSGOTLNQVSVSPRPQLO 2285
Db 1880 KPSPSDNLISAFSTSDGAISVPSLSAP---GQGTSTNTV--GATVNS-----Q 1923
QY 2286 IOQPOP 2291
Db 1924 AAQAP 1929

RESULT 6
US-09-854-856-14
; Sequence 14, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walker, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19

; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 2136
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)..(2136)
; OTHER INFORMATION: xaa - Any Amino Acid
US-09-854-856-14

Query Match 2.5%; Score 364.5; DB 4; Length 2136;
Best Local Similarity 18.1%; Pred. No. 4.7e-13;
Matches 414; Conservative 298; Mismatches 725; Indels 849; Gaps 96;

QY 399 TNKAGSNKSFLLAANEELIESIRAKKGDDIDNVKSPETEKKNETENDSKDAENREF 458
Db 160 TSKDRPVSPQSLVSGKEEPPARSGGG-----SAKEPQERSQOODDI-----EEL 207
QY 459 EQDSLEKSDDDKTPDDDPQKSEVGDFFKSEKSNGLSESPGAGKAGSSTRI--ITRLRN 517
Db 208 ETKAVGMSNDGRFLKFDIEIGR--GSFXT-----VYKGLDTETTVEVAVCEL 252
QY 518 PDSKLSQLKSQVAAAHEANKLFEGKEVLVNVNSOGEISRLSTKKEVIMKN--INNIF 575
Db 253 QDRKLTKSRQR-----FKBAEMLKGLQHPNIVRFYDSWESTVKGKCIVLVT 301
QY 576 KLGQEGKYRVHYNOYSTNSFALNKHQHRD-----HDKRRHLAHK-----FCLTPA 621
Db 302 ELMTSTLTATYTKRFKVMKIKVLKRWCRQILKGLQFLHTPTPIHRDLKCDNIFITGPT 361
QY 622 GEFKWNQSVHGSVKVLTISTRLTITOLENNIPSPFLHPNWNASHANWIKAV----- 672
Db 362 GSVKIGD-----LGLATLK-----RASFAKSVLGTPEFMAP 392
QY 673 QMCSKPRE-----FALALAILECA-----VKPV-----VMLPIW 701
Db 393 EMVEEKYDESDVYAFGMCMLMATSEYPYSECQNAAIYRRVTSVKKPAFDKVAPEV 452
QY 702 RE-----FLGHTRLHRMTST-----EREKEK-----VKKK 727
Db 453 KELIEGCIQNDKDERYSIKDLLNHAFFQETGVRVELAEEDDCEKTAIKLWLRIEDIKKL 512
QY 728 EKKQEEETQOATWVKYTFPVKQVWQKQGEERYVTGY---GGWSWISKTHYVREVPKL 784
Db 513 KGKYKDNEA-----IEFSFDLERDVPDVAQEMVESGVYCEGDHKTAKA----- 557
QY 785 PGNTVNYRKSLGCTNNNDENMDSDKSKSRSPAKIKIEPDSEKDEYKGSDAAGADQ 844
Db 558 -----IKDRVSLIKRQORQLVREPEKK-----KOEESLKQOQVEQSASQ----- 600
QY 845 NEMDISKITEKKDQDVKELLDSDSKPCKEPEMEVDMDMKTESHVNCQESSQVDVYVNSE 904
Db 601 -----TGIKQLPSASTGIP-----TASTTSASVSTQVEPEPEA 634
QY 905 GFHLRTSYKKTKSSKLDGLLEKRIKQFTLEEKQRLKLEGGIGKIGCTSTNSKNLS 964
Db 635 DQHQQLQYOQPSISVLSLSDGTVD-----SGQSSSVFTESRVSS 671
QY 965 ESPVITKAKEGQSD-----SMRQEQSPNANDQPEDLIQCSQSDSSVLRM 1011
Db 672 QQTV-----SYGSOHQAHSTGTVPGHIPSTVQASQPHGYVPSSVAQSQSQPSSSL 727
QY 1012 SDPSHTNKLKPKDRVLDVSRSPETKCPKQNSIENDI-----EEKVSDLASRGQEP 1064
Db 728 TGVSSSQPTQHPQQO-----QGIIQOTAPQOTVQVLSQTSSEATTAQPVSPQAP 780
QY 1065 T-----KSKTKGNDFFIDDSKLASADDIGTLCKNKKPLQESDITVSSSKSALHS 1116
Db 781 QVLPQVVSAGKQSTQG-----VSQVAPAEFVAV-----AQPTTTASSVSDSAHS 827
QY 1117 SVPKSTNDRDATPLSRAMDFEGK-----LGCDSSENSTL 1150

QY 2607 CKTPYDESKFYIGCDRCQWYHRCVGILOSEAELIDEYVCPQOSTEDAMTVLTPLTEK 2666
Db 2306 LET-----QIEC-----LMSDQECVK-----RAREEEIEQLNEVIEK 2337
QY 2667 DYGLKRV--LRSLQAKHMAWPLEPDPNDADPYGVYVIKEPMDLATMEERVQRRYYEKL 2724
Db 2338 LQELANIGOKTSMNAHLS-----EADS-----LKHQLDVIAEKLALEQOVETA 2384
QY 2725 TEVADMTKIFDNCRYNPNPSDFYQCAEVLESFFVQKLGKFGKASRHNKLOS 2778
Db 2385 NEEMTEMKNVKETF-----KMQLTQELFSLKRERESVEKIQS 2424

RESULT 5
US-09-854-856-46
; Sequence 46, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 2076
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2076)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-46

Query Match 2.5%; Score 364.5; DB 4; Length 2076;

Best Local Similarity 18.1%; Pred. No. 4.5e-13;
Matches 414; Conservative 298; Mismatches 725; Indels 849; Gaps 96;

QY 399 TNKARGSNKFLAAANEIELESIRAKKGOTIDNVKSPETEKKDNETENDSKDAEKNREEF 458
Db 100 TSKDRPVSPSLVSGSKKEPPPARSGGG-----SAKEPQERSQQDDI-----EEL 147
QY 459 EDQSLKSDDDKTPDDPEQKSEVGFKEKXNGELSPGAKGASGSTRITRLRN 517
Db 148 ETRAVGMSNDGRFLKFDIEIGR---GSFKT-----VYKGLDITETTVAVANCEL 192
QY 518 PDSKLSQLKSOQVAAAHAANKLFECKEVLVNVNQSQETSLSTKKEVIMKGN--INNYF 575
Db 193 QDRKLTKEQR-----FKEEAEMKLGQHPNIVRFYDSWESTVKGKCKIVLVT 241
QY 576 KLQGEQKIRYVHNOYSTNSFALNKHQHRED-----HDKRRHLAHK-----FCLTPA 621
Db 242 ELMTSGTLKYLKRFVKMKTIVLRWCROILKGLQFLHTPTPIIHRDLKCDNIFITGPT 301
QY 622 GEFKNGSVHGSVLTISTLRLTITQLENNIPSSFLHPNWSHRANWIKAV----- 672
Db 302 GSVKIGD-----LGLATLK-----RASFAKSVIGTPEFMAP 332
QY 673 QMSCKPRE-----FALALATLECA-----VKPV-----VMLPIW 701
Db 333 EMEEYKESVDVYAFGCMLEMATSEYPYSECQAAQIYRRVTSYGVKSPASFDKVAIPEV 392
QY 702 RE-----FLGHLRHRMTSI-----EREEXE-----VKKK 727
Db 393 KEIEGCIRQNKDERYSIKDLLNHAFFQETGVRVELAEDDGEKTAIKLWLRIEDIKKL 452

QY 728 EKKOBEETMOQATWVKYTPPVKHQVWKQKGEYRVITY---CGWSHISKTHYRVFVKPL 784
Db 453 KGKYKONEA-----IEFSPDLERDVPEDVAQEMVSGYVCEGDHKTMAKA----- 497
QY 785 PGNTNINYSKLBGTNNMNDENDESKRKCSPPKKIKLEPDEKDEKGVGSDAAKGAQD 844
Db 498 -----TKDRVSLKRRKREQQLVREGEKK-----KQEESSLKOOVESSASQ----- 540
QY 845 NEMDISKITEKKDDVDKELDSDSKPCKEEPMEDDDMKTESHVNCQESSQVDVNVNSE 904
Db 541 -----TGIKQLPSSASTGTP-----TASTTSASVSTOVEPEPEA 574
QY 905 GFHLRTSYKKTKSSKLDGLLERRIKQFTLEEKORLEKIKLEGIGKIGTSTINSSKNLS 964
Db 575 DQHQQLQYQQPSSISVLSDGTVD-----SQGSSVFTESRVSS 611
QY 965 ESPVITKAKGECQSD-----SMROBQSPNANNDDQEDLITQCGSDSSVLRM 1011
Db 612 QQTV-----SYGSHQEAHSTGTVPGHIPSTVQAQSPGHVYPPSSVAQGSQGSQPSSSL 667
QY 1012 SDPSHTTNKLYPKDRVLDVDSIRSPETKCPKONSINDI-----EEKVSDLASRQEP 1064
Db 668 TGVSSSQPIQHQQQ-----QGIQQTAPPQTVQVLSQTSSTSSSEATTAQPVSQOPAP 720
QY 1065 T-----KSKYKGNDFIDDSKLASADDIGLICNKKAPLQEBSDTIVSSKSKALHS 1116
Db 721 QVLPQVSAGKQSTQG-----VSQVAPAEPAV-----AQP--QATQPTTFLASSVDSAH 767
QY 1117 SVPKSTNDRDATPLSRAMDFEGK-----LGCDSSENSTL 1150
Db 768 DVASGMSDGNENVPSSGRHEGRTTKRHYKRSVRSRHEKTSRPKRLILNVNKGDRVV 827
QY 1151 ENSSDT-----DSIQ-----DMIVQNSNESISEQFRTREQDVEYLEP 1192
Db 828 ECQLETHNRKMTFKFDLDGDNPEEJATIMVNDFFILAIERESFVDQVR---ETIEKADE 884
QY 1193 LKCELVSGETGNCEDRLPVKGTANGKKPSQOKKLEER---PVNKCSDQIKLKNITDKK 1249
Db 885 MLSEDSVSVEPEGD--QGLSLQGDYGFSGSKLEGEFKQPIPASSMPOQIGIPT----- 938
QY 1250 NENRESEKKGQRTSTFOINGKDNKPKIYLKGE--CLKEISESRVSVGNVEPKVNNINKI 1307
Db 939 -----SSLTQV-----VHSAGRRFVSPVPSRRLRESKVFPP----- 969
QY 1308 IPENDIKSLTVKESAIRPFINGDVIMBDFNERNSETKSHLLSSSDAEG-----NYROS 1361
Db 970 ---SEITDTVAASTAQSPGN---LSHSASSLSLQQAFLSRAQMTGEGPNTAPPNFSHT 1023
QY 1362 LETLPSTK-----ESDSTQTTTPSACPSBSNVQVEDMEIETSEVKKVTSPITSEESNL 1418
Db 1024 GPTFPVVPPLSSIAGVPTTAAATAPVPATSPSPND--ISTG---VIOSEVTVPTEGI 1077
QY 1419 ---SNDFTDENGFLPINKNENGVESKRKTIVITEVTMTSTVATESKTVIKVEKGDQK 1473
Db 1078 AGVATSGVTVSGLPI---PPVSESPVLSVVSSIIIPAVVSIISTSPSLQPTSTSEI 1134
QY 1474 VVSTENCASKSTVTTTTTKLSTPSTGSGVDIISVKEQSKTVVITVTTDSLTGTGTL 1533
Db 1135 VVSSSTALYPSVTSATSA-----SAGGSTATPGKPPA--VVSQQAAGS--TTVGATL 1183
QY 1534 VTSMTVSKEYSTRDVKVLMKFSRPKTRSCATALPSYKKFVTKSTKKSIFVLNDDLLKLA 1593
Db 1184 TSVSTTTTSPSTASQLSI-----QLSSSTSTPTLAETVVSVAH----- 1221
QY 1594 RKGIREVPYFNNAKPAIDIMVPPSPRPTFGITWRYRLQTVKSL--AGVSLMLRLWLAS 1651
Db 1222 -----SLDKTSHS---TTGLAFSLASFSSSSSPGAGVSSIIIS----- 1256
QY 1652 LRMDMAAKVPPGGGSTRTETSETITTTTIIKRRDVGPIRFEYICIRKIICPIGPVET 1711
Db 1257 -----QPGG-----LHPLVIPS 1269

Db 529 ---GLILEKCALQROLEDVEEL-----SFSREQ--IQARQOTIAEQ-SKLNHAHS 576
QY 685 LAILECAKPVNLPWIFREFLGTHLRMTS-----IEREEKEKVKKKKEEETMOQA 740
Db 577 LSTVEDLKAEIVASSESREKELEKHEAVNTYKIKLEMLEKNAVLDRMAESQAEUER 636
QY 741 TWVKYTFPVKHQWQKGB--EYRVTYGGGWSWISKTHVYFVPKLPONTNVNRYKSLE 797
Db 637 LKQTLFHSHEELSCLKDELEHRIN-----IEKLDNLGIHYKQOID 680
QY 798 GTKNNDENMD--ESDKRCSRPKKIKTEPSEKDEKVGSDAAGAD-----QNEM 847
Db 681 GLQNEMSQKTIETMQEKNLITKQNLILEISKLQLOSLVNSKSEEMTLQINELQKEI 740
QY 848 DISKITEK--DQDVKEI-----LQSDSDKCKEPEMEV 879
Db 741 ELTROBEKEKLEGOEVOELQKTELLEKQMEKENDLOEKFAQLEAE--NSILKDEKKT 799
QY 880 DDDMKTESHVNCQESSQVDVNVVSEGFHLRTSYKKKTKSKLDGLLERRIKOFTLEEKOR 939
Db 800 EDMLKHTPVQOE-----RLFLDSIKSKSDSVWEKEI-EILIEENED 843
QY 940 LEK--IKLEGGKIGKGTNSKNL-----SESPVITKAKECCQSDSMRQE-----Q 985
Db 844 LKQOCIQLENEIEKORNTSFSAEKNFEVNYQLOEYACLLKVKDDLEDKKNKQOELEYS 903
QY 986 SPNANNDQEDLIQCSQSDSVLRMSDPSTHTNKLKPKDRVLDVYSIRSEP----- 1038
Db 904 KUKALNEE-----LHLQRINPTVTKM--KSSVFDEDKTFAETLEMGV 946
QY 1039 -----RCPKQNSIENDIEKVSCLASRGQEPTRKSKTGNDFFIDDSKLASADDI 1087
Db 947 EXDTTELMEKLEVTKREKE--LSQRLSDLS-----EQLKQKHGEISFLNEEVKSLKQE-- 998
QY 1088 GTLICKNKPLOEESDITVSSSKSALHSSVPKSTNDRDATPLSRAMDEFGKLGDSSEN 1147
Db 999 -----KEQVSLRCRELEIIINHRA-----ENVQSCDTQVS 1029
QY 1148 STLENSDPVSIQDSEEDIMVQNSNESISEOFRTREQDVEVLEPLKCLVSGESTGNCE 1207
Db 1030 SLL-----DGVVMTSRGAGSVSKVNSKSGESKINVEDKVFE-----NMTVGESKQEQ 1081
QY 1208 ---DRLP--VKGTEANGKPKPOOKLEERPVNKCSQIKLNTDKNNENRESEKKGOR 1262
Db 1082 LILDLHLPVTKESSLRATQPSNDKLQ-----KELNVKSEQNDLR-LQMEAGR 1129
QY 1263 ---TSTQI--NGDNKPKIYLKGECLKEISESRVWSGVNPEKVNINKIIP 1309
Db 1130 ICLSLVYSTHVDQVREYMEKND-----KALCSLKEELIFA--QEEKIKELQK-IH 1177
QY 1310 ENDIKSLTVKESAIRPFING-----VIMEDFNERNSETSKSHLLSSSDAEGNYRDSLE 1363
Db 1178 QLELOPMKQET-----GDEGKPLHLIGLKQAVSECSYFLQTLCSVLGEY----- 1225
QY 1364 TLPSTKESDSTQTTTPSASCPSNSVQVEDMEIETSEVKKVTSSPITSEESNLSNDFI 1423
Db 1226 -----YTPALKC-----EYNAEDKEN-SGDYI 1246
QY 1424 DENGLPINKN--ENVNGESKRKTIVET-----TMTSTVATESKTVIKVEKGD 1470
Db 1247 SENEDPELOQRYEVODFOENMHTLLNKVTEENYKLLVLQTRLSKWTGQOTOMKLEFGE 1306
QY 1471 KQTVSSSTENCAKSTVTTTTVTKLSTPSTGSDVSIISVKEQSKVTVTVTDTSLTTTG 1530
Db 1307 -----ENLPKE-----ETEFLSHQ----- 1322
QY 1531 GTLVTSMTVSKYSTKDKVLMKFSRPPKTRSGTALPSYRKKFVTKTKKSIFFVLPNDLKL 1590
Db 1323 -----MTNLEDIDVNHKSKLSLODLKTEKLEEQVQLESLSISSLQOQ-----LK 1367
QY 1591 KLARKGGIREVPVFNNAKPALDIWYFPPRPTFGITWYRLQTVKSLAGVSLMLRLWA 1650
Db 1368 ETEQ-----NYAEB-----IHCLO--KRLOQVNS----- 1388
QY 1651 SLRWDDMAKVPVGGGSTRTETSETETITTEIILKRDDVGPYGIREFYCIKIRKICPGVPE 1710
Db 1389 -----ESTVP-----SLPVDVSVLITESDAQRTMYPGS-----CVKNI--DGTIE 1427
QY 1711 TPKETPTPORKGLRSSALRKPETPKQGPVLIETWA--EBELEWEITRAFAERVEKKB 1769
Db 1428 FSGEFGVBETNIVKLEKQYQOELEEVAKVIVSMSIAFAAQOTELSRGSGKENTASSK 1487
QY 1770 -AAVVEQOAKKLEQOKPVIATSTSTSSITSPISPAKVWVAPISGVTTGKMWLT 1828
Db 1488 QAHAVCOQBQHYFNEMK-----LSQDQIGFQFETVDVKFKEEFKPLSKELGEHGEKILL 1542
QY 1829 TKVGPATVTFQONKNFHOTFATWYKQGSNGVQVQOKVILGIIPSTSTGTSTQOFTSFQ 1888
Db 1543 S-----NSDPHD-----IPESKDCVLT-----SEEMFSKDK 1569
QY 1889 PRATVTVIRPNTSGSGGTTNSQVITGPQIRPGMTVIRTPLOOSTLIGKAIIRTPVWVQPG 1948
Db 1570 TFIVRQSIHDEISVSSMDASRQMLNNEOELE-----MRQELVRYQ 1611
QY 1949 APQVMTQIIRQGPVSTAVSAPNTVSTPGQKSLTSATSTSIQSSASOPPRPQOGVKL 2008
Db 1612 EHQQA-----TORSSIDNENLVSERERVILLELEALKQLSLAGREKL 1653
QY 2009 TMAQLTQLTQGHGNGQLTWTVIGQGQQTGTQ-----LQIPQGV--TVLPQGGQOLMOA 2060
Db 1654 CCELRNSSTQONG-----ENQGEVEEQTFKEKELDKPDPVPEILSNERYALOKA 1706
QY 2061 AMPNGTVORFLTPTPLATTATTA-----TTTTVSTTAAGTGEQKQSKLSPOMQVH 2111
Db 1707 --NNRLKILLVVKVTTAAVEETIGHVLGILDRSSKQSSASLITWRSEAEASVKSCVH 1763
QY 2112 QDKTLP-----PAQSSSVGP-----AKAQPTAOPSARPOPTQOPSPAQOE----- 2153
Db 1764 EEHRTVDSIGISYSGSDMPRNDINMWSKVTEGTELSORLVSRSGFAGTEIDPNEELML 1823
QY 2154 -VQTOPEVOTQTTVSSHVPSEAPTHAQSKSPQVAAQS-----QPOSNVQOGSPVRVQSP 2207
Db 1824 NISSRLQAAVEKLELAISETSSQLEHAKVTQTELMRESFROKQEATESLKCOBELERLH 1883
QY 2208 SQTIRPSTPSQLSPQ-----QSQVQTTT-----SQPIQHTSLQIIFS 2248
Db 1884 EESRAEQVAVELSKAEVGVIDYADKTLFERIOIEKTDIDIRLEQELLICASNRLQELBA 1943
QY 2249 QGQOPQOPQVQSTQTLSSGQTLNVQSVSPSPQLIQOPQOPQVIAVPLQLOQVQVLSQ 2308
Db 1944 E-----QQQIQBERELLSR-----QKEAMKAEGP-----VEQ-----QLLQETEKLMK 1982
QY 2309 IQSQVVAQTOAQQSGVQPOQIK-LQLPIQIQSSAVQTHQIQNV-----VTVQAASVQEQLO 2363
Db 1983 EKLEVQCAEKVRDLDLQKVAKALEIDVEPQVSRFIEQEKNTLMDLRQOQNALEKOLE 2042
QY 2364 RVQOLDOQKKQOQOIEIKREHTLQASNOSELIQK-QVVMKHNA-----VIEHLKOKK- 2416
Db 2043 KMKRFLDEQDAIDREHERDVFOEQIKLEQOLKVVPRFPQISEHOTREVQOLANHLKEKT 2102
QY 2417 --SMTPAEREEENORMVCMVKYIILDKID--KEEKQA-----AKKKREESVE 2461
Db 2103 KCSSELLSKFQLOLDI--QERNEETIEKLEFRVRELEQALLVSADTFQKVEDRKHFGAVE 2159
QY 2462 QK-----RSKONATKLSALLPKHKEQURAEILKKRALLDK--DLQIEVQ 2503
Db 2160 AKPELSLEVQLOAERDAIDRKEKEITNLEEQ-----EOPFRELENNEEVQQLHMOLEIQ 2215
QY 2504 -----BELKRDIAKKEKDLMOQAQATAVAAPCPPTVPLPAPPAPPPPPPPGVQ 2555
Db 2216 KKESTTRLEQLEQENKLFKD-DMEKLGLA-----SSSSKKKKMISTTSKTKKDTLYCI 2245
QY 2556 HTGLLSTPLTPVASQK-----RKREBEKD-----SSSSKKKKMISTTSKTKKDTLYCI 2606
Db 2246 ESDAMSTQDQHVLFGRFAQIQEKEVEIDQNLNEQVTKLOQLKITTNDKNVIEKNEILRD 2305

Db	1650	STSESDSASTSLSDSTSTSVSESTSTSTSVSSASNSTSTSLSD--DGRSTSLSDSTST	1707
Qy	2057	LMQAAMPNGVQRFLFTPLATTATTTAAAGTGEORQSRLPQMGMHQDKTL	2116
Db	1708	STSESSTSTSES--DSDSASTSLSESTSTISDSTSTSDSASTSMVSVDNRASTSL	1765
Qy	2117	PPAQSSSV--GPAKAQPOTAPQARPPOPQTQPSAQPEYQTOPEYQTOTTVSSHVPSEA	2174
Db	1766	SDSTSTSVSDSTSASTSASEASTSTRESESTASTLSESTSTSVSDSTSTSTD---SAS	1822
Qy	2175	OPTHAASSKPDVAAGQPOGNSVQGSPVRVQSPQTRIRPTSPQLSPGQQSOVQTTSQ	2234
Db	1823	TSTSESDSNSESTLSGSTS-----TSVSDSTSASTSASASTSTVSDSNSASTSLSGST	1877
Qy	2235	PIPIQHTLIQIPSQGPQSQPOQVQSGSTOTLLSGQTLNOGVSSVSPSRPOLIQOQPQVI	2294
Db	1878	STSVSDSTSTSTSASTSTSESDSDSASTSLSGSTSTSLSDSTSTSTSDSASTSTSE--	1935
Qy	2295	AVPQLOQQOVVLQIOSOVYAQIAQAQSGVPPQIKQLPIQIOQSSAVQTHQIVVTWQ	2354
Db	1936	-----SASTSTSVSESDSESTSVSESSS-----TSVSDSTSTSTSESASTSTSE	1979
Qy	2355	AASVQEQLRVQQLRQOQKKKQOOEIKREHFLQASNQSEIILQKVMMKHNAVIEHLKQ	2414
Db	1980	SESTSES-----TSVSESSSTSI SDSSTSTSMSTSETFTTSQPSINSEQIGDSLSE	2032
Qy	2415	KKSMTPAERENORMVICNOVMKYILDKIDKEKQAAKKRKEESVEQKRKNQATKLSA	2474
Db	2033	DTIVT---QSKNTNML--NKTKG----DYDLQEGRGYTDEQHNEVTSQNQADNHNNLDL	2083
Qy	2475	LLFKHKBLRAETLK 2490	
Db	2084	L---HONRLQDKVVVK 2096	

RESULT 3

```

RES001.3
US-09-134-001C-5080
? Sequence 5080, Application US/09134001C
? Patent No. 6380370
? GENERAL INFORMATION:
? APPLICANT: Lynn Doucette-Stamm et al
? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
? TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
? FILE REFERENCE: GTC-007
? CURRENT APPLICATION NUMBER: US/09/134.001C
? CURRENT FILING DATE: 1998-08-13
? PRIOR APPLICATION NUMBER: US 60/064,964
? PRIOR FILING DATE: 1997-11-08
? PRIOR APPLICATION NUMBER: US 60/055,779
? PRIOR FILING DATE: 1997-08-14
? NUMBER OF SEQ ID NOS: 5674
? SEQ ID NO 5080
? LENGTH: 3696
? TYPE: PRT
? ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

```

Query Match	2.98	Score 410	DB 4	Length 3696
Best Local Similarity	17.78	Pred. No. 1.9e-13		
Matches 464	Conservative 427	Mismatches 1027	Indels 704	Gaps 108
Qy	311	QKNKPYIRHPIGYD-RSRKRYWFLNRR--LIITEEDTENENEKKIWIYSTKVQLAELIDC	367	
Db	569	QVNKDII---PSNTLASYNKYNKIKERAQTVLDEETNTP-----	606	
Qy	368	LDKDYWEAEALCKILEENREETHRHMDITEDLTNKARG-----SNKSFLAAANEELIESI	421	
Db	607	FNQRYSQTDIDLLHQLTILNRYVSASREINDKAQEMTDVAVDSTELTTTEKDTLVQDI	666	
Qy	422	RAKGGDIDNVKSPE-----ETEKDKNETENSDKAENREFEEDOSLE	464	
Db	667	ENHKNETSNIDDELTDGQGVVERKEAGLHLESDTPHPVTKPNARQVNNRAD-QQRTLI	725	

Qy	465	K0SDDKTPDDDB-----QKSEVGDGFKSEKSNGELSESQAGKAGSGSTPIITR	514
Db	726	RNNHEATTEQNEAIRQVEAHSSDALAKGEAETDITVNEARD-----NGTKLIAYD	777
Qy	515	LRNPDSKLSQKSOQVAAAHAEANKLFKKGKELVVVNSOGEISRLS-----TKKEVI	566
Db	778	VPNPTK-----KAEARAAVTNSANSIKD-----INNNTQATDERNDALVNRSKDEAI	828
Qy	567	MKGNINNYFKLQGEGRYRVVHNOYSTNSF-----ALNKHQHREDHDKRRLAH	614
Db	829	Q--NINT--AOGNDDVTEAQN--GTNTIQOVPPLTPVKRONAIATINAKADEQKRLIOAN	882
Qy	615	KFCLTPA--GEFKWNGSV-HGSKVITITSLRITITOLENNITPSSFLHPNASHRAWNIK	670
Db	883	NNATTEEKADAERKVEAVITANQNTNATTNTRDVEDQAQTT-----GSGIIS	929
Qy	671	AVOMCSKPREFALALALECAVKPVVMLPIWREFLGHTRLHRMTSTEREKEKVKKKEKK	730
Db	930	ATSPATKIKEDARA-AVEAKAIAQNOOI-----NSNNMATTEEDKDALNQEVAHK	978
Qy	731	QEEETMQQATWVKVTFPVKHQWVKQGBEYRVTYGCGWSWISKTHYRVFPKPLGNTNV	790
Db	979	QAIAIATINQA-----QSTQQVSE-----AKNNGINTI	1005
Qy	791	NYRKSLEGTKNNDENMDSDKKCRSPKKIKIEPDSEKDEVKGSAAKAGDONEMDIS	850
Db	1006	NODQPNVAKNNTKTILEOKGNKKS-----AIAQTDPDATTEE-----KQBAVS	1049
Qy	851	KITEKKDQDVKELLSDSDKPCKEEPMVEDDDMKTESHCQESSQVDDVVNVSEGPHLRT	910
Db	1050	AVSQAVTNGITHINQANSNDVDQOE-LSNAEQIITHNVNVQKKPQ-----ARQALIAKT	1103
Qy	911	SYRKKTKSSKLDGLERRIKQF-TLEEKQRL--EKI-----KLEGGIKIGKYS	956
Db	1104	NEQSAINSNDEGTIEEKOKAIOQLNDAKNLADEQITQAASNQNVDMALNIGISNISKIO	1163
Qy	957	TNSKNLSESPVITKAKGECQSDSMRQEQSPNANDQPEDLIOGCSQSDSVLRMSDPHS	1016
Db	1164	TNFTK--KQARQDVQNOKPEKEAELNSTPHATQDEKQDALTRLQAKETALNDINQAO	1220
Qy	1017	TTNKLYPKDRLV-----DDVSIRSPETKCPKQNSIENDTEEVSDLSARGQEPPTS	1067
Db	1221	TNQNV---DTALTSGTQNTQNTQNVYRK---KQEAKTINDIVQHQKQSONNDDATTEE	1274
Qy	1068	KTRGNDFFIDDSKLASADDITGLICKNNKPLIQEESTIVSSKSAIHSSVPKSTNDRDA	1127
Db	1275	KEVANN-----LVNASQ-ONVYSIKDNATTNNQIDGIVSDGRSQSINATPDTSIKRNA	1326
Qy	1128	TPLSRAMDFEGLGCDSESNSTLENSDTSIQ---DSSSEDMIVONSNESISEQFRTRE	1184
Db	1327	-----KNDIDIKAADPKTKIQINDATDEE--IQEANKKIEE-----	1361
Qy	1185	QDVEVLBPLKCELVSGESTGCEBRLPVKTEANGK----KPSQOKKLEERPV--NKCS	1238
Db	1362	-----AKIEAKDNIQNRSTRDQVNEAKTNGINKTIENTITPATTVKSEARQAVONKANE	1413
Qy	1239	QIK-LKNTTDKKNENRESEKKQORTST-----FQINGKDNPKPI-YLKGECLKELS--	1288
Db	1414	QINHIQTPDATNEEKQEA--NVRVSAELARVQAQINAEHTTQGVKTIKDDAITSLIRS	1470
Qy	1289	-----ESRVVSGNVPEKPVNNIKIIPENDIKSLTVKESATRPFGINGDVIMEDFNERNSS	1343
Db	1471	NAQVVEKESARNAIEQATQOTQOTFINNNDNATDEEVA-----NNLVIA---TKQKSLD	1522
Qy	1344	TKSHLLSSDAE-----GNVYRDSLETLPTSKESDSTOTTPTSASCPESNSNQVEDMEIET	1399
Db	1523	NINLSNNDVENAKVAGINAEIANVLPAATAVAKSKAKKIDQKL---AQOINQIOTQOTAT	1579
Qy	1400	SEVKVTSPTSEENLNDNFIDENGLPIKNNENNVNGSEKRTVITETVTMTSVATE	1459
Db	1580	TEEK-----EAAIQLANQKSNEARTAIQNEHSNNGVAQAK-----	1614

US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 3.0%; Score 432.5; DB 4; Length 2137;
Best Local Similarity 17.2%; Pred. No. 3.8e-17;
Matches 395; Conservative 355; Mismatches 999; Indels 547; Gaps 77;

QY 394 ITEDLTNKGARNSKFLAAANEET-----LESIRAKGDDIDNVKSPETEKD- 440
DB 149 MSSEVTNDQSEKAGISQSSSETSNOSSELNTYASTDHSVETTTNNDNTAQODNKNSSN 208
QY 441 --KNETENDSKDAEKNEEFEDQLEKSDDKTPDDPEQCKSVGDFKSEKSGELSES 498
DB 209 VTSKSTQSTSSSEKNTSSNLTQIETKATDSLATSAPRSTNOISNLTSTSTNQSSPT 268
QY 499 PGACKGASGPRITIRLNP-----DSKLSQLKSOQVAAAHEANKLF-----KEG 544
DB 269 SFANLRTFSRFTLVNTMAAPTSTTTTSSLTSSVVVKNDFNEHMLNLSGATYDPKGT 328
QY 545 KEVL--VVNSQGEIS---RLSTKEVIMKGNIN--NYF----- 575
DB 329 IATLTPDAYSKGALSINLRDLSNRFFFTKVLGNRYEGYSPDGVGGDGFAGFSPG 388
QY 576 ---KLQEG-----KYRVVHNOYNSPALNKHORE-----D 605
DB 389 PLGOIGREGAAGVIGGLNNAFGKLDYHNTSTPKSDAKADPRNVGGGAFGAFVSTD 448
QY 606 HD-----KRRHLAHKFLTPAGE-----FKWNGSVHGSKVLTISTRLTIITOLENNIPSS 655
DB 449 RNMATTEASSAAKLNVOPTDNSQDFVIDYNGD---TKVMTVYAGQTFR----- 497
QY 656 FLHPNASHRANWIKAVQMCSPREFALALAILCAVAPVVMPLIWRFLGHTRLHRMWS 715
DB 498 -----NLTDWIKN---SGGTTFLSMTASTGGAKNLQOQVFGTFEYTESAVAKVRY 545
QY 716 IEREE-KEKVKKKKEKQEEETMQOATWVKYTFPVKHQV--WKQKGEERYVT-----GY 766
DB 546 VDANTGKDILPPKTIAGEVDATN-----IDKQLNLNKNSGSYSTDALQNSNY 595
QY 767 GGWSNI-----SKTHVRYFVKPLPGNTNVMYRKSLGKTKNNMDNMOESDKRKCIRS 818
DB 596 SETSGTPTLKLTNSSQTVIYKF-----KDVQGPQISVD-----SQPREVGKT 637
QY 819 PKKIKI-EPDSEKD-----EVKGSAAKAGADQ-----NEMISKITEKKDQDVRELL 864
DB 638 INPITITTDNSKVDLTTVTGLPSGLSFQDQTTNTITGTPSEVGTITVT-----VNTT 690
QY 865 DSDSDKCKEPMEDVDDMKTESHVNCQESSOV-----DVNVNS 903
DB 691 DATGNVTSKQFTIITQITISPVNVNTPSQASEVTPINPITAITADNSGKVVHTVITGLP 750
QY 904 EGFHLRTSYKKTKSSKLDGLLERIKOFTLEEKORLEKIKLEGKIGKIGKISTNSSKNL 963
DB 751 QGLKFDASTNIVGTPPTQIGTNTITIESTDASGNKTTTKINYE-----VTRNSADS 802

QY 964 SESPVITKAKEGCO-----SDSMRQEQSPNANNOPEDLIOGCSQSDSVLRMSDPSTHT 1018
DB 803 TSTIVNSVSTSIENSTSLSDSVKASQSLSTSKLSLSE--SASTSNTSIQASASASTS 860
QY 1019 NKLYPKDRVLDDVSIKSPETKCPKQNSIENDIEEKVSDLASRQBPETKSKTKGNDFFDD 1078
DB 861 KQL-----SEASTSTSDASASAKSESTSKASTS-----LSE 893
QY 1079 SKLASADDIGTLICKNKKPLIQEESDITVSSSKSALHSSVPKSTNDRDATPLSRAMDFEG 1138
DB 894 STSTSVSDSASVSTSES---ASTSTSVSGSTSTSTSTSTSDSASIKASESAS--TS 948
QY 1139 KLGCDSSNSTLENSSDTIVSODS--SEEDMIVQNSNISSEOFRTREDQVLEPLKCEL 1197
DB 949 KLLSESVSTSTSDASTSTSVSDNSASTSKSTSTSTSTSTSD----- 997
QY 1198 VSGESTGNCEDRLPVKGTGANGKPKSQOKKLEERPVNKCSQOIKLKNNTDKKNENRESE 1257
DB 998 --SASTSTSE-----SESDASTSLESTSTSTSVSDSTS--TSTSDSASMSASE 1043
QY 1258 KKGQRTSTFQINGKDNPKIYLKGECLKEISESRVVSQNVGPKVNNINKIIPENDIKSLT 1317
DB 1044 SNSKSTS-----LSE-STSTSLSGSTSEAST-----SDSASTS 1074
QY 1318 VKESAIRPFINGDIVMEDFERNSESTKSHLLSSDAEGNYRDSLETLPSTKESDSTQTT 1377
DB 1075 TSESE-----SDSTSTSLSESTSTSLSGS--TSASTSDSAST--TSTSESDSTSES 1120
QY 1378 TPSASCPESNVQVEDMEIETSEVKVKTSPITSEESN-----LSNDFIDENGL 1428
DB 1121 T---SLESSTSVSDSTSEASTSE---SASTSTSESNASTSLSGSLSTSDSTST 1173
QY 1429 PINKNNVN--GESKRKVTITEVTTMTSTVATESKTVIKVEKGDQTVVVSSTENCASKPV 1487
DB 1174 STDSASTSESESDSTSTSLSESTSTSLSDSTSTSEASASTSESDSTSESTSLSE 1233
QY 1488 TTTTIVTKLSTPSTGGSVDI---ISVKBQSTV---TTVTVDSTLTGGLTVMVTS 1540
DB 1234 STSTSVSDSTSDASTSTSVSDSEASTSLESTSVSDS--TSTSDSASTSTST 1292
QY 1541 KEYSTRDKVKLMKFSRPKTKRGTALPSRYKRVFKTKSKTSFVLPNDLKLARKGGIRE 1600
DB 1293 ESDSTSESTSLSESTSVSDSTSA--STDSASTSTSES-----ESDASTSLSG---- 1341
QY 1601 VPYFNYNAPALDIWPPSPRPTFGITWRYRLQTVKSLAGVSLMLRLWLASLRWDDMAAK 1660
DB 1342 -----STSTSLS----- 1348
QY 1661 VPPGGGSTRTETSETETTTTEIIKKRRDVGPGIRFEYCIRKIIICPIGVPTPKETPTPQR 1720
DB 1349 -----DSTSTSDSASTSTST-----ESDSEASTSL 1375
QY 1721 KGLRSSALRPRKRPETPKQTPGV-----IETWVAEELELWEIRAFAPER 1764
DB 1376 SGSTSTSLSDSTSTSDSASTSTSVSDNSASTSLSGSLSTSVSDSTSTSDSAST 1435
QY 1765 VEKEKAQAVEQAKRRLBQOKPTVIATSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1824
DB 1436 SEDS-----SEASTSLSGSTSTSTSDSTSTSTSDASTSTSVSE-----SNSTSTSI 1484
QY 1825 MVLTATKVGSPATVTFQKNKNEHOTFATWVKQGSNGSVVQVQOKVGLIIPSTGTSOOTF 1884
DB 1485 ESLSTSV--SDSTST-----STSDSASTSTSVSDSDASTSTSSSESV-----STSDSE 1532
QY 1885 TSFQPRATVRI-----RPNTSGGGTTSNQVITGPQIRPGMTVIRTPLOQSTLGRKAI 1939
DB 1533 TSTSDSASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSDSAS---ASTSESDSDAST 1589
QY 1940 RTPVMVQPGAPQVMTQIIRQOPVSTAVSAPNTVSSSTPGQK---SLTSATSTSIQSSAS 1996
DB 1590 SSSESSTSVSDSTSEASTSEASTSTSVSDNSASTSLSESTSTSLSDSTSMSTSDAST 1649
QY 1997 QPPRQOQGVKLTMAQLTQLTQGHGNGGLIVVQGGQQTTCGQLQLIPQGVTLVLPGQQQ 2056

RESULT 2

ALIGNMENTS

RESULT, T 1

THIS PAGE BLANK (USPTO)

Db 12354 QEVSVKETPTDISDRKPKTEKATSIILSEQESISVQEVSVKKEAPGCTVETAKPKTEQATSI 12413
Qy 109 AIYEVLRNFTGTVLRSPFREDCAALYSQEQCT-LMAEMHVLLKAVLREEDTSNTFTG 167
Db 12414 SPHEISIQEISVKEAP---TDICDKPKSEKATSIILSEQESIAVQEVSVKE-----A 12463
Qy 168 PADLKDSVNTLFLFDGWTWPEVLRVYVCESDKKEYHHVLPYQEAEDYPYGPVENKIKVLQF 227
Db 12464 PGSTEEAKPT-----EYAKSSISPHESLTVOEV----- 12492
Qy 228 LVDOFLTNTIAREBELMSEGVIOYDDHCRVCHKGLDILCCETCSAVYHLECVKPP--LEEV 285
Db 12493 -----LYKEASAEISDKKPKQTKATSIILSEQESIAVQEVSVKKEAPGRVEDV 12538
Qy 286 -PEDEWQECVCAHVKGVTDCV-----AEIQNKPYIRHEPIGYDRSRKRYWFLNRR-- 337
Db 12539 KPKTEQATSIILSEQESLTVOEVTVKGDSEISDKKP-----KTEKAISILSEQS 12588
Qy 338 LIIEEDTENENKKIWIYSTKVQLAELIDCLDKDYWEAELCKILEEMREIHRHMDITED 397
Db 12589 ISVOEITVKEAPGSKVEDKLPQA-----TSVISPLESLTVQEVSVKSESGD 12636
Qy 398 LTNKARGSNKSLAAANEIILE---SIRAKKGIDINVKSPEE-----TEK 439
Db 12637 ISDRRPKTERATSIIFSQESISVQEVSVKKEAPGTADVDPKPEEQATSIILSEQESLTVOEV 12696
Qy 440 DKNETENDSKDAKNREE-----FEDQSL-----KSDDKTDPDDP---EQGKS----- 481
Db 12697 TVKEIPTDISDKKPKSEKATSIILSEQESISVHEVSVKDPAGSMKDAKPKTEQATSVLSPH 12756
Qy 482 -----EVGDF--KSEKSNGLS-----ESPAGKAGSGSTRI 511
Db 12757 ESITVQEVSVREPTIELSDKKPKSEKATSIILSEQESISVQEVSVKKEAPGSMKDAKQTEQ 12816
Qy 512 ITRLRNPDLSQLSKSOQVAAAHEANKLFKEGKVELVNVANSQGEISRLSKKEVIMK--- 568
Db 12817 ATSVISPHESLTVOEVSVKKEAPTELSDRKPKKEKATSIILSEQESIS---VQEVSVKEAP 12872
Qy 569 GNTNNYKLGQEGKYRY--HNOYSTNSFALNKHQHRHEDHKKRHLAHKCLPAPAGEFKW 626
Db 12873 GSKVD-LKLKTEQATSVISPHESLTVOEVSV-KEAPTEISDKK-----PKTE--- 12917
Qy 627 NGSVHGSKVL-----TISTLTLTIT---QLENNIPGSFLPNWAS--HRANWIKAVQCMCK 677
Db 12918 ----QATSIILSEQESISQEVSVKAGPSLKDSEKPKSEQATSIILSEQESLTVOEVTVKEA 12973
Qy 678 PREFALALALECAVKPVVPLPIWREPLGHTRLHRTSIEREBE-----KVKK----- 726
Db 12974 PADIS-----DVKE-----KTEKATSIILSEQESISVQEVSVKKEAPGSM 13011
Qy 727 KKKQEBEE-----TMQOATWVKYTFPVKQHVQKQGEYRVTVGYGWSWISKT 775
Db 13012 KDAKPKTEQATSVISPHESLTVOEVSV-VK---EVPTEISDKKPKSEKATSIILSEQESISV 13067
Qy 776 HVYRFYKPLPGNTNVNRYKSLGTX-----NNMDENDESDKKRCSRPKKIKLEPD 827
Db 13068 HEVS-VKESPGSMKDAKPKTEQATSVISPHESLTVOEVSVKESPTIELSDKKPKKEKATSI 13126
Qy 828 -SEKDEVKGS-DAAKGAQDNEMDISKITEKKDQDVKELL-DSDDKPKKEEPEVEVDMD- 883
Db 13127 LSEQESISIQEVSVKEVPESMKDLKPKTEQATSVISELLSILSVQEVSVKESPGDLSDKP 13186
Qy 884 KTE---SHVNCQESSQDVVNVSEGPHLRTSYKKTK--SSKLDGLLERIKOFTLEEKQ 938
Db 13187 KTEKATSIILSEQESISVQEVSVKKEAPGSEEVKPKTEQATSIISPHMSLTVOEVTVKE-- 13244
Qy 939 RLEKILEGGIKIGKTSTNSKNLSRPVITRAKECCQSDSMRQESPNANNDQPDLLI 998
Db 13245 -----VPADISDVKPKSEKATSIILSE-----QESISVQEVSVKKEAPGSMK----- 13285
Qy 999 QGCSQSDSVLRMSDPSTHTNKLKPYKDRVLDDVYSIRSPETKCPKQNSIENDIEEKSDLA 1058
Db 13286 -----AKPKTEQATSVISPHESLTVOEVSVREVTEISDKK 13321

Qy 1059 SRQEPFKSKTKGNDFIDDSKLASADDIGTLICKNKKPLIQEESDITVSSKS-----A 1113
Db 13322 PKSEKATSIILSEQESISVQEVSVKKEAPE---SMKDAKPK-TEQATSVISPHESLTVOEV 13376
Qy 1114 LHSSVPKSTNDROATPLSRAMDFEGLCGDSSESNSTLSESSDVTVSIODSSEEDIVQNSN 1173
Db 13377 SVKEVPTESDKKP-----KTEKATSIILSEQESISVQEVSV-----VKEAP 13416
Qy 1174 ESISEQFRTREQDVEVLEPKCELVSGESTGNCEDRLPVKGTTEANGKKPQOQKKL---E 1229
Db 13417 GSVKDLKPKTEQATSVISP-----HESLTVOEVSVKKEAPTEISDKKPKTEQATSVLSE 13469
Qy 1230 ERPVNCSDQIKUKANTTDKKNNENRESEK-----KQORTSTPOINCK-----DNKPK 1276
Db 13470 QESIS--IQEISVKEAPGSLKDEKPKSEQATSIILSEQESLTVOEVTVKEAPADISDVKPK 13527
Qy 1277 IYLKGECLKE-----ISESRVSVGNVEPKVNNIKIIPENDIKSLTVKESATRPF 1326
Db 13528 TEKATSIILSEQESISVQEVSVKKEAPGSMKDAKPKTEQATSVISPH-----SLTVQEVSVR-- 13583
Qy 1327 INGDVIMEDFNERNSSSETKSHLLSSSDAEGNYRDSLETLPSTKESDSTQT--TTPSASCP 1384
Db 13584 ---EVPTEISDKKPKSEKATSIILSEQESISVQEVSVKKEAPGSMKDAKPKTGOATSVLSPH 13640
Qy 1385 ESNSVNOVEDMEIETS-----EVKVTSPITSEESNLSNDFIDENGLPINKNENVNGES 1440
Db 13641 ESITVQEVSVKKEAPTELSDRPREKATSIILSEQES-ISVQEVSVKKEAPESMKKEAKPRTE 13699
Qy 1441 KRKTIVTEVMTSTVATESKTIVTKVEKQKQTVVSVSTENCAKSTVTTTTTTVTKLSTPS 1500
Db 13700 QATSVISELLSL-----SVQEVSVKSEPEFSDKKPKTERATSIILSEQESISVQEVSVKE 13754
Qy 1501 TGGSDVIISVK-EQSKTVVTTTVDLSLTGGLT-----VTSM-----T 1538
Db 13755 AGSVEEVKPKTEQATSIISPHM--SLTVQEVTVKGPADISDVKPKSEKATSIILSEQES 13812
Qy 1539 VS-KEYSTRQVKVLMKFSRPPKTRSGTALPSYRKFTVKTSKSFIVLPNDLKLAKRGG 1597
Db 13813 ISVQEVSVKKEAPGSMKDAKPK-KTEQATSV-----ISPHESL--TVQEVSV 13853
Qy 1598 IREVPYNYNAKPAALDIPYSPRPRTGIIWRYRLQTVKSLAGVSLMLRLMLWASLRWDDM 1657
Db 13854 VREVPTEISDKKP-----KTEKATSIILSEQ-----ESISVQEI 13886
Qy 1658 AAKVPPGG---GSTRTETSETTEITTEIKRRDVGPIRGIFEYCIK----- 1701
Db 13887 SVKEAPGSMKDAKQKTEQATSVISPHESLTVOEVSVKEVPTESISDKKPKSEKATSIILSEQ 13946
Qy 1702 ---IICPIGVPETPK-----ETPTPQKGLRS--SALRPKRPETPK 1737
Db 13947 ESISVQEVSVKKEAPESMKDAKPKTEQATSVISPHESLTVOEVSVKELPTIELSDKKPKTEK 14006
Qy 1738 QTPGVILETVWAEELWEIRAFARVEKEQAQVQQAQKRLQKQKPTVIATSTTSP- 1796
Db 14007 ATS-ILSE---QESISVQEVSV-----KEAPGSMKDAKPKTEQ-----ATSVLSPH 14048
Qy 1797 -----TSSTTISPAQYVNAVAPISGVSVTGTGKWLVTTKVGS 1833
Db 14049 ESLTVQEVSVKEVPTDISETKPKTERATSIILSEQESISIQEVSVKEATG-----SMKDAK 14103
Qy 1834 PATVTFQONKNFQTFATWVKQGSNGVVOQKVLGIIPSTGCTSQOFTTSQPTAT 1893
Db 14104 P-----KTEQATSVISELLS--LTQEVSVKESPEFSDRKKPKTET 14142
Qy 1894 VTIRPNTSGSGGTTNSNQVITGPIQIRPGMTVIRTPLOQSTLGRKAIL--RTPVMVQCAPQ 1951
Db 14143 AT-----SILSEQESISVQEVSVKKEAPCSMEEVKPKTEQAT---TIIISQESLTVOEVAVK 14195
Qy 1952 QVMTQIIRGQP---VSTAVSAPNTVSTPQOKSLTSATSTSNIOSSASQPPRQOQGVKL 2008
Db 14196 EVPTDISDKKPKQOQATSVFSP-----QESISVQEVSVKKEAPGAVKDSKPKPKTEHAKS 14247

```
Db 1513 TEKTVE---ADDKFTTVPVLAGDEEESNLPKLPQDIFEEEA-PVAVTTAAPSKDD---GE 1565
QY 1440 SK-----RKTVITEVMTSTVATSKTKVIVKEGDKQKQVVSSTENCAKSTVT 1487
Db 1566 QRPVEVEEKPIEDGQKPIEDEFTPTPS-----SENEIEPESDRATTIASKEE--PSEPS 1618
QY 1488 TTTTITTKLSTPTSGSDVSIISVKEQSKVTVTTTVDLSLTTRGGTLVTSMTVSKEYSTRD 1547
Db 1619 TGAPTCKDEPAEPSTDAPEDES-KETPESEVPTTAPA---GEKIPTSSITPDEEPTAT 1673
QY 1548 KYVLMKFSRPPKTRSGTALPSYRKFKVTKSKSIFVLPNDLKLARKGGIREVPYFNVN 1607
Db 1674 SAPVAKPDEVEDKETSTEIPTDAPASSEDESN---STDQIPS-----EVP-----E 1717
QY 1608 AKPALDIWYPSPRPFTGITWRVLOTVKSLAGVSLMLRLWLASLRWDDMAAKVPBG-- 1665
Db 1718 KRP-----EVLAEIPOPSTETGI 1768
QY 1724 RSALRPKRPKQGPVVIETWVAEELELWEIRAFARVEK--EKAQAVEQ-----Q 1776
Db 1769 KOQDETTAAPSIDRK-----EPYVTEIDEEATTVAPISEKDEKPTBEBKPVQKPTGEE 1822
QY 1777 AKRLQEQKPTVIATSTTSPTSTSTISPAOKVMVAPISGVSVTTCTKMLVLTIKVGSPT 1836
Db 1823 PSEEEKRPIEDVSTEGVPVSEASEAGSTES-----SEEVKPSTEGEVAEKPRD--- 1873
QY 1837 VTFQKNKHFOTFATWVKGQGSNS-GVVQVQKVLGIIPSSTG---TSQQTFTSFOPRTA 1892
Db 1874 -----KQPSSTAQAPVETIPEISTELPAQDGDKPTSEAPVDSDEDTSA 1916
QY 1893 TVTIR-PNTSGSGCTNSQVITGPOIRPGMTVIRPLOQSLGKAIITPVMVQPGAPQ 1951
Db 1917 PSDEKIPVSG-----EEVEGPEV-----TTASPAQAAEDELKTPAESEPSSTD 1960
QY 1952 QV-MTOILIRGQVSTAVSAPNTVSTPGQKSLTSATST-----SNTQSSASOPPRPOQO 2005
Db 1961 KVPETEYQKPEDETKADEPESVTVQSDVATSTAPVAGDDEKDKQATASP--EEEEE 2018
QY 2006 VKLTMAQLTQLTQGHGNGGLTVVIOGQQTGQLQIIPQGVTVLP--GPGQOLMO-AAM 2062
Db 2019 IKPTIAPAAEIPO-----PSEKEPVDE-QEVESGTKATPAESDGPIDETIAPA 2065
QY 2063 PNCVTQREFLTPLATTATTASTTTTSTVSTAAGTGEQORQKSLSPQMVHODK-----TL 2116
Db 2066 TSGPIDE-----ASTAAPTKESTTVASAA-----SP--AVHDDKIKDVTTTQ 2106
QY 2117 PPAQSSSVGPAKAQPOTA-----QPSAR-POPQTOPQAPQPEVOTQPEVOT---QTTV 2166
Db 2107 PVADEKEVAAPQDETSTSDVSTDSFTAQDDEKQDKTEAPVAPTTVSSPTADSAASSIP 2166
QY 2167 SSHVPS-----EAQPTH---AQSSKPOVA-AQSQPQSNVQGSQSPVRV--QSPSQTRTRPS- 2215
Db 2167 TVEVPSPVEIDTRKPMDDIMSQTITAPHTADGAASTSTEDQAPVTVSPQDAEKTPVSPAP 2226
QY 2216 -----TPSQLSQGOQSVQVTTTSQ-----PIPIQPTHSLQIPSGQPOQPOVQSSQTQ 2264
Db 2227 QSDSKTPSEAPQADAEIPATATPLDNNKIPATVAPQTDGVPATAPALDEDKIQITAAP 2286
QY 2265 L-----SSQQTUNQVSVSPSRPOLIQOQPQVIAVLPQLQOQVQLVSLQIQSQVVAQIQ 2318
Db 2287 LDEEKIPSTAAPLDDDEKIPAPVSPVFDVEPSSEKPAVSEYDGE-----ESTEPVHDVE 2341
QY 2319 AQSGVPOQIKQLPLTQIQOSSAVQTHQIONVVTVQAA-----SVOEQLRVQOQLRDOQO 2373
Db 2342 TSTDEPTSDAKLPPTSPATPSPSEPAETAEIYPETAAPLEKEVPEKATEQPELEKETP 2401
QY 2374 KKQOQLEIKREHTLOANSQSEI-----IQOVVMKHNVAIEHLKOKKSMTP 2420
Db 2402 EKATEQPELEKETPEKATEQPELEKETPEKATEQPELEKETPEKATEQPELEKETPE 2461
```

```
QY 2421 AER-----EENORMIVCNQVMKYILDKIDKBEKQAAKKRKRREESVEQRRSKQNA TKLSA 2474
Db 2462 VVKPSLDSTEEDE-----ESVESEESADKKDNKETEEDTDKKHEBEVPA 2508
QY 2475 L-----LFKHKEOLRAEILKKRALLDKDLQIEVQEEELKRD LKTKK 2516
Db 2509 VVSEIQPQSEEA VPTTGHPLFPH---LASSTTTPPAVDDR-----VGEE-----DEEN 2553
QY 2517 DLMOLAQAATAVAAPCPVPT---PVLPAAPPSPPPPPGVQHTGLLSTPTLPVVASQXR 2572
Db 2554 TTVKLSSSTTTSTTESPVTSAPSTTTVASQQOQPIPTPPYG--HA-----PEYEDY 2603
QY 2573 KREEEKDSSSKKKMISTTKETKKDKTKLYCICKTPTYDESKFYICGDCRCQNMWYHGRV 2632
Db 2604 DEEEVPGTCRYAGKLYVSAQOIPRDDPCDFCF-----PFSDDII 2644
QY 2633 GILQS-----EALIDEYVCPCQ-----STEDAMTVITP 2662
Db 2645 CLQQSCPPPIAGCHEBPISGFCPRVECPVSMAAVLNITTTTSTTTLPP 2695

RESULT 15
Q95YM2 PRELIMINARY; PRT; 17352 AA.
AC Q95YM2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE I-connectin.
OS Procamburus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidea; Cambaridae; Procamburus.
ON NCBI_TaxID=6728;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21423462; PubMed-11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RL EMBO J. 20:4826-4835(2001).
CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB055861; BAB64297.1; -.
DR InterPro: IPR000577; EGGY kin.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; ig; 48.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; Igc2; 13.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; 3.
DR PROSITE; PS00835; IG_LIKE; 50.
DR PROSITE; PS50002; SH3; 1.
KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 4.0%; Score 572; DB 5; length 17352;
Best Local Similarity 17.7%; Pred. No. 4.1e-15;
Matches 576; Conservative 570; Mismatches 132; Indels 874; Gaps 133;

QY 3 SEEEEDGDAETQDSEDEDEMEEDDDSDSYPEMEDDDDDASYCT-----PSSFR 56
Db 12294 STEISDKPKSEKATSI LSEQISVQKDPAGSIKDAKPKTEQATSIISPESLTI 12353
QY 57 SHSTYSSTP---GRRPRVHRPRSPILEE-----KDIPLPFFPKSSDLMPVNEHIMNVI 108
```

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Posler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Godek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003750; AAF56376.1; -
DR FlyBase: FBgn0039257; CG13648.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF00093; vwc; 1.
DR SMART: SM00214; VWC; 4.
DR PROSITE: PS01208; VWF; 1.
SQ SEQUENCE 2768 AA; 294032 MW; CA929A21774E4684.CRC64;

Query Match 4.08; Score 572.5; DB 5; Length 2768;
Best Local Similarity 18.4%; Pred. No. 4.1e-16;
Matches 562; Conservative 437; Mismatches 1147; Indels 905; Gaps 128;

QY 78 PILEKDIP---PLERPKEEDLMVNEHMVIAEVLNFGTVLRLSPFRFEDCAA 134
DB 84 PIITCPVPVDVPHSPPECTELSIPEKF---GCSTIEEKYPGAGVPSNPKPCELCYC 140
QY 135 LVSQEQCTLMAEMHVLLKAVLREEDTSNTTGPADLKDSVNSTLYFIDGMTWPELVRY 194
DB 141 INNOTKVMQ-----ENELD-----FMDQSTTTTTTTRPTTGFILASTMT 209
QY 195 CESDKYHVLVPOEADYDYPGVENKIKVLQFLVQDQFLTN-----TCSAVYHL 275
DB 164 NKGS-----CCPVRYSCDH-----ENELD-----FMDQSTTTTTTTRPTTGFILASTMT 209
QY 237 -----TARELMSEGV-IQYDDHCRVCHKL-GDLLC-----CE-----TCSAVYHL 275
DB 210 PTTTDCIHDGEIFADGASLKGKACNCHCYCMRGDIVCAVOCEVPMMAANGKSCRAMPA 269
QY 276 ECVKPPLEEYPEDEWOCVCAHKVPGVTDV-----AEIOKNKPYIRHEPIGYDRSR 329
DB 270 EGECCPSNYCEDDSSSTEIVETTPESATSVPAKHAIPEKREDVDLQEHIDDDKNKE 329
QY 330 KYWFLNRLIIEEDTENENKKIYWTSTKVQLAELIDCLDKDYW---EAECLKILEMRE 386
DB 330 TATIPSAELSGTEIEEKEKD---KATTVPVQVTD--EKDFSEFEPESSTAGIPSDSR- 383
QY 387 EIRHMDITDLTKARGSNKSLAANEEILESIRAKK-----GDDINVK----- 432
DB 384 -----IDLPSSTESKESSTERAEDDIVKIVTTPPEGSGEEDVPKPSQIPEKEIT 434
QY 433 -----SPEE-----TEKDKNET-----ENDSKD 450
DB 435 EDELKIVTSAPAKASPEEEVVTATTSAPTEEDVKPTTACTISEEEEGKPTAEESG 494
QY 451 AEKN-----REFEQQ-----SLEKSD-----DKTPDDDPQOKSEY 483
DB 495 EEKDVKVTAAPEETEAEKPTAPVASDEKEQPKPSGSGDELDLKPPTAPTAGATSA 554
QY 484 GDFKSEKSNELSESP-----GAGKASGSTRIITRLRPDPSKLSQLKSOQVA 531

DB 555 SESEEDDEKSTEARPTSVDDIEPAKPTSESEASGEGEDVAKETTPAGEASTAGEEIV 614
QY 532 AAAHEANKLFKEGKVLV-----VNSQGEISRLST-----KKEIVMK 568
DB 615 KGTTPAGEPSGEGDEIVKGTTPAEBSSESEDELTKVTTPAGEPSVAGEEIAKETTPA 674
QY 569 GNTNNYFKLGQBGKYRYVHNOYSTNSFALKNHQRHDKRRHLAHKFCULTPAGEFKWNG 628
DB 675 GEPS---IAGEEEIVK-----TTPAGESIIAGEEEIVK-----VTTPAGESSSEG 717
QY 629 SVHGSKVLT-----JSTLRLTITOLENNITPSSFLHPNASHR 665
DB 718 EEEIIKVTTPAGESSEGDDEEIVKESTPAGEIPISEGEEDVIKATTSAPKSDIEGVKEPET 777
QY 666 ANMIKA--VOMCSKPREFALALALECAVPKPVMLPIWREFLGHTRHLRMTSIEREEKEK 723
DB 778 ATEVPAEEVEDFAKP-----TTPIAEAEPIAGTPIPTDGIS-----GEEIVK 822
QY 724 VKKKEKKQBEETMOQATWVYTFPVKHQVWKQGEERYVTGYGGWSWISKTHVYRVPK 763
DB 823 GTTTPQTLLEEQPEISEESTEV---PVAED-----D 848
QY 784 LPGNTVNV-YRKSLEGTKNMNMDESDRKRCSRSPKKIKIEPDSKED----- 831
DB 849 LSSSTSASAIASSTEGVQDDAASETSSAPARAGDKDEAATVTPAQDKDDEEVEDATDL 908
QY 832 ---EVKGSAAKAGADQONEMDISKITEKKDQDVKELLDSDSKPCKEEPMVEVDDMK --- 884
DB 909 PVEDVQSTTAKTTTTEQPKAESSTEADAEIEVTTSSPADKQEVPAEPADKDKDEED 968
QY 885 -----TESHVNCQESSQVDVNVVSEGFHL-RTSYKKKTKSSKLDGL 924
DB 969 VQTATLPIKSDIGPPVVDTEATTGQPETSETATDKPPSVYLPVPVQSEVPSSTAKDNR 1028
QY 925 LERRIKOFTL-----EKKORLEKIKLEGIGK-----IGKSTNSKSLSESPVITKA 972
DB 1029 NDETETKPTLPPSGEDOSSEPLPAMDLPAGIPGEGDCLVEGKYANNITVPATPCOVSC 1088
QY 973 KEGCQSDSMROBQSPNANDPOE-----DLIOGCSQS-----DSSLRMSDPSHTNKLYP 1023
DB 1089 K--CISLVAQCQWECKLPENLEKCTVAADLLDGCCTYICDESTESAEDSTAK--P 1144
QY 1024 KDRVLDVSTRSPETKCPKONSNDIEEK-----VSDLASRGQEPKSKTK 1070
DB 1145 DNKIDEDVSEISTE-EIPKDVIMPTGITEQPLSHVKPDEBIEQIPVTSYPAQFDESTAKVD 1203
QY 1071 GNDFFIDDSKLASADDIGTLICKNKKPLIOEEDTIVSSSKSALHSSVPKSTNDRDATPL 1130
DB 1204 KKP--IDES-----AEDKRFIGSEED-----SKPIDESEEDKKPV 1237
QY 1131 SRAMDFEGKLGCDSESNSTLSESDTYSIODSSEEDMIV-----QNSNESISEQPRTR 1183
DB 1238 E-----ESAEDKKPVDESEKEKPLPTVPASEIEKESKPEDEKKT 1278
QY 1184 EQD-----VEVLEPLKCELVSGSTGNCEDRLPVKGTANGK---KPSQKKLEERPVNKC 1236
DB 1279 EADFAAPTEQPEATTPAIAADTAKEKVEDDKLATTAPVSGEDELKPADEKKRTE----- 1332
QY 1237 SDQI---KLANTTDKKNENR---ESEKKGORTST--FOINGKDNKPKIYLGECLEKI 1287
DB 1333 TAQIPDAEIPASTDEPESSTELPTVDLDKKPEEDSTKGTAPESDKVPEVPTASTENEI 1392
QY 1288 SESRVSGNVPEKPVNNINKIIP--ENDIKSLTVKE-----SAIRPFINGDVI----- 1332
DB 1393 EESDKFTTVAPPKISASDETEPTAEEDLVPAETPEIESEFEVSTKKFAVQGPPLTLAPA 1452
QY 1333 -MEDFNERNSESTKSHLLSSDABGNVRSLETLPSTKESDSTQTTTP-SASCPESNSVN 1390
DB 1453 QPEKKPVDAETADISTEAPSAEVEKASGETSESDNEIDAGASSTPVVPSADEDKTSP 1512
QY 1391 QVEDMELETSEVKKVTSSPIT-SEESN---LSNDFIDENGLPI-----NKNENYNGE 1439

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celisner S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome.";
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RA EMBL; AE003808; AAM70936.1; -;
DR EMBL; FBgn0013988; Strn-Mick.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; Ig; 29.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00408; IGC2; 29.
DR SMART; SM00406; IGV; 4.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00835; IG_LIKE; 26.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Immunoglobulin domain; Transferase.
SQ SEQUENCE 9270 AA; 1036749 MW; 0DD82D040B27C2AD CRC64;

Query Match 4.1%; Score 580.5; DB 5; Length 9270;
Best Local Similarity 18.2%; Pred. No. 8.2e-16;
Matches 574; Conservative 533; Mismatches 1160; Indels 885; Gaps 133;

QY 4 EEEE-----EEDGDAETQSDSEDEDEM-----EEDDDSDYPEEMDDDDASVCT- 51
DB 2614 EEELKASAKKQGDQDIEQKQSEVSEVVAEIKSEGIIEPKKPEEM-DTEAKSEKATV 2672
QY 52 -----ESSFRSHSYSTPGRRKPRV-----HPRSP----- 78
DB 2673 LDKQVLEKELEASAEKQGDQDVEKKSQKPEVSEVVAEKISEETIEPKKPEVKTDEIKS 2732
QY 79 -----ILEEKDIPPLEFPKSSDLMVPNE--HIMNVIAIYELVNRFGTVLRSLSPFR 127
DB 2733 EKATALDKQVLEKELEASAEKQGDQDVEKKSQKPEVSEVVAEIKSEGIIEPKKPEEM-DTEAKSEKATV 2781
QY 128 FEDFCAALYSQEOCTLMAEMHHVLLKAVLREE--DTSNTTFFGADLKDSVNSLYLIDGM 185
DB 2782 IEPKPEVKTDE--TKSEKATALDKQVLEKELEASAEKQGDQDVEKKSQKPEVSEVVAEIKSE 2831
QY 186 TWPEVLRVYCESDEKHYHVLVYQAEADYVPGVYENKIKVLQFLVDFLATNTAREELMSE 245
DB 2832 --FEVSEVVAEKISE-----ETIEPKKPEVKTDE-----KSEKATVLDK 2870
QY 246 GVLYQDDHRCVCHKLGLDLCCTCFCSAVYHLECVKPLLEVPPEVDEWQCEV--CVAHVPGV 303
DB 2871 QVLEKELEASAEKQGDQDVEKKSQKPEVSEVVAEIKSEGIIEPKKPEEM-DTEAKSEKATV 307
QY 304 TDCVAEIQNKPIRHEPIGYDRSRKRYWFLNRLIIEEDTE-----NENEKKIWIYSTKV 359

DB 2911 T-IEE--PKKPEVKTDEIKSEKATA-----LDKQVLEKELEASAEKQGDQDVEKKSQK 2962
QY 360 QLAELIDCLDKYWEALCKILEEMREIHRMD-ITEDLTNKARGSNKSPFLAANBEIL 418
DB 2963 EVSEVI-----AE--KISEEKIEPKKPEKETEVEKSEKATVLDKQVL-----BEKEL 3008
QY 419 ESTRAKGDI-----NVKSPETEKKKNETENESDKAEKAREE 457
DB 3009 EASAKQGDQDVEKRSQKPEVSEVVAEIKSEGIIEPKKPEVKTDEAKSE--KATTLQVQV 3067
QY 458 FEDQSEKSDDKTPDQDPEQ--GKSEVGDPKSEKSNSELSSESPGACKGASGSTRITRL 515
DB 3068 LEERELEA--SAQKQGDQDVEKKSQKPEVSEVVAEIKSEKIEEP----- 3110
QY 516 RNPDSKLSQKSOQAAAAHEANKLFKEGKEV-LVVSQSG--EISRLSTKKEV----- 565
DB 3111 KKPEKETEVEKSEK-----ATVLDKQVLEKELEASAEKQGDQDVEKKSQKPEVSEVVAEK 3166
QY 566 IMGNINNYPKL-----GOGEKYR-----VYH 587
DB 3167 VSEGIIEPKKPEKETEVEKSEKATTLDKQVLEKELEASAEKQGDQDGRSDDIITLK 3226
QY 588 NOYSTNSFALNK-----HOHRE-----DHDK-----RRHLAKHFCILTAPGEKFN 627
DB 3227 ERUTELSKALGSSVDEILRESREIVNLEDDKVVAKHLFKLRDHIVHIVTGKKEENK-- 3284
QY 628 GSVHGSKVLTISTRLTITOLENNIPSSFLH-----PNWASH--RANWIKAVQCMCKPRE 680
DB 3285 -----EXELPESFIELLECEASPEAAEKVKLVLAIEK-----TN 3318
QY 681 FALALAILECA-----VKPVVPLPIWREPLGHTLRHMTSISEREKEKVKKK---EKQ 731
DB 3319 VILTKATIQLIDDSNMFTKPSLLIP-----KLENLER--VAVKIQSETYVDKSS 3365
QY 732 EEBETHQOATWVKT-----FPVKHGVKOKGEYRVTYGGHGSWISKTHYRVFVPLPG 786
DB 3366 EKMSISLQOQLMDFVILDDFLDDETEVLKPKIENIKITLLSDYIDIEKKDGPLLTAVING 3425
QY 787 NTWV---NYRKSLEGTKNNDENNDSDRKRSRPPKKIKIE--PDSEKDEKVGSDA--- 838
DB 3426 KINWVSQHILTIIEEVK-QLTENHDOKEKDVSNAAEDNFADEKREESQKEIKDSEAKHK 3484
QY 839 -AKGADONEMDISKITEKKDQDVKELLDSD----- 869
DB 3485 KSKVSEKKSIEEKLDEKKEKQTESAIDEKSKAEVSEIYSEKITDEKAEKQKVEYKGS 3544
QY 870 -----KPKCEEPMEVDMDKMTESHVNCQESSQVDVYN-VSEGE---HLRYSYK 913
DB 3545 EAKPKKAKVLEKKSIEEKLDEKKEKQTESAID-EKSKAEVSEIYSEKITDEKAEKQK 3603
QY 914 KTKSSKLDGLLERRIKQFTLEEKQRLKIKLEGGIKGIGKTTSTNSKNLSSESPVITKAK 973
DB 3604 KEVKDS-----EAKPKKAKVLEKKSIEEAKLE-----DKKETQDSDAIDEKSKAEVS 3651
QY 974 EGCQSDSMRQEQSPNANDQPEDLIQCGSQSDSVLRMSDPSTHTNKLIPKDRVLDVSI 1033
DB 3652 E-TVSEKITDEKAEKQKEVKD-----SEAKPK---KAKVLEKKS 3689
QY 1034 RSPETKCPKONSNDIEEK-----VSDLAS-----RGOEPTKSKTGNDFFIDSKLA 1082
DB 3690 EEEKLDEKKEKQTESAIDEKSKAEVSEIYSEKITDEKAEKQKVEKQSEAKPKKAKVL 3749
QY 1083 SADDIGTLICKNKKPLIQEESDTIV--SSSKSALHSSVPKSTNDROATPLSRAMDFEGL 1140
DB 3750 EKKSIEEKLDEK-----ETQDSDAIDEKSKAEVSEIYSEKITD-----EKAQESQKEE 3800
QY 1141 GDSSESNSTLENSDVTQSDSEEDMIVONSNESEIQEPTREQDQVEVLEPKCELVSG 1200
DB 3801 VKDSEAKPKKAKVLEKKSIEEKLDEKKEKQTESAIDE-----KSKAEVSE-IVSEKITD 3855
QY 1201 ESTGNCEDLRPLVKGTEANGKKPSQOKLEERPV--NKCSQDIKLKWT---TDKKNENRE 1255
DB 3856 EKAQESQK--EVKGSSEA---KPKKAKVLEKKSIEEKLDEK--KEKQTESAIDEKSKAEV 3910

Db 4390 -----EAKPKKAKVLEKKSIEBEKLEBKKEKQTESAIDEKSKAEVSEIVSEKI 4438
QY 1879 TSQOTFTSQPRTATVTRPNTSGSGGTTNSOVITGPIQIRPGMTVIRTPLOQSTLGRAI 1938
Db 4439 TDEK-----AQSKEEVKDEAKPKKAKV---LEKKSIEAK 4473
QY 1939 IRTPVMPGAPQVMTQIIRGPVSTAVSAPNTVSTPGKSLTSATSTNSIOSSASOP 1998
Db 4474 LEDKRETQDSDAIDKSO---KAEVSEIVS---EKITDEKAQES-----QKEVVKDEAKP 4523
QY 1999 PRPQO-----GOVKLTMAQLTOLTOGHGNGOGLTVITOGQGTGQIQLIPQGVTVLP 2052
Db 4524 KKAKVLEKKSIEBEKLEBKKEKQTESAIDEKSKAEVSEIVSEKIT--DE 4571
QY 2053 PQQLMQAMPNGTVQRELTPLATTATTSTTTTSTVTAAGTGEQROSKLSPOMOVHQ 2112
Db 4572 KAQSKEVVD-----SEAKPKKAKVLEKKSIEBEKLEBKKEKQTESAI 4616
QY 2113 DKTLPAPSSVGPAPAKAPQTAQPS-----ARPOPTQPPQAPQEVQTOPEVQTO 2163
Db 4617 DEKSKAEVSEIVSEKITDEKAQESQMEVVDSEAKPKKAKVLEKKSIEBEKLEBKKEKQTESA 4676
QY 2164 T-----TVSSHVPSEAPHAOSSKPOVAQSOQSNVQOQSPVRVOSQRTI 2212
Db 4677 TDSAIDEKSKAEVSEIVSEKITDEKAQESQKEVVDSEAK-----PKKAKV 4723
QY 2213 ---RSTPSOLSPOGOOSQVQTT---TSQPIQIPTSLSIQPSGQPOQO-POVQSSSTOTL 2265
Db 4724 LEKKSIEBEKLEBKKEKQTESAIDEKSKAEVSEIVSEKITDEKAQESQKEVVDSEAKP 4783
QY 2266 SSGQTLNVQSVSSRPQIQOQPOQVITAVPOLQOQVOVLQSOIQSVVQAQOAOOS--- 2322
Db 4784 KKAKVLEKKSII---BEKLEBKKEKQTESAIDEKSKAEVSEIVSEKITDEKAQESQK 4839
QY 2323 -----GYP-----QOIKLOLPIQIOQSSAV---QTHQIONVTVVQAASVQEO 2361
Db 4840 EYKDEAKPKKAKVLEKKSIEBEKLEBKKEKQTESAIDEKSKAEVSEIVSEKITDEKAQ 4899
QY 2362 LORVOQLRQOQKKQOQOQ-----ETKREHTLOASQSEIIOQVMKHNATIEHLQ 2414
Db 4900 ESQKEVVDSEAKPKKAKVLEKKSIEBEKLEBKKEKQTESAIDEK---FQKAEVSETVSE 4956
QY 2415 KKSMTPAEENORMIVCNQVMK-----YILDK--IDKEKQAARKKRREESVEQKRK-- 2466
Db 4957 K--ITDEKAESRKEEVVDSEAKPKKAKVLEKKSIEBEKLEBKKEKQTESAIDEKSKAE 5014
QY 2467 -----QNAKTLKLSALLFKHKEQLRAEILKRLALDKLQIEVQOELKRDLLTKKEDLMQ 2520
Db 5015 VSETVSEKITDEKAQESQKEVVDSEAKPKKAKVLEKKSIEBEKLEBKKEKQTESAIDEK 5066
QY 2521 LAQATAVAAPCPVTPVLPAPPAPPPPPPPGVQHTGLLSPTLTPVASQKRKRBEKDS 2580
Db 5067 --TETKVATDKSQT-----VEVSEIVLEKISEEKAESQKVELKDS 5106
QY 2581 SKSKKKKMMSTSTKTKDKTKLYCICKTPYDESIFYIGDCRCQNWYHGRVCVGLIOSEAE 2640
Db 5107 EAKSKKAKVLE-----KKST-----LKEKLENDKKQKEDGATN-----RSQKAE 5146
QY 2641 LIDEVVCPOQSTEDAMTVLTPLETKDYGLKRLVLSQAHKMWPFLPEVDP----- 2692
Db 5147 AAD--VVPKISEEKAESQKVELKDS-----EPDSKAKSKPD 5178
QY 2693 --PNADPYGYVKEPMDLATMEERVQRRYRYEKLTEFVADMTKIFDNCNRYNPSPPFYQ 2750
Db 5179 GLPADEKSHGAKVSESVVPKNEAKTDQLSAKKPTVLDLTV-----PKRKP-VL 5228
QY 2751 CAEVLESFFVQKLKGF-----KASRSNNK 2775
Db 5229 AEQTADSILOTYKSMDSYKDKESRSARK 5260

RESULT 13
Q8MLD9

ID Q8MLD9 PRELIMINARY: PRT: 9270 AA.
AC Q8MLD9;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG18255-PA.
GN STRN-MECK OR CG8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7272;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fouts K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu K., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski R.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE FROM N.A.

Query Match		4.1%; Score 580.5; DB 5; Length 7210;
Best Local Similarity		18.2%; Pred. No. 6e-16;
Matches		574; Conservative 533; Mismatches 1160; Indels 885; Gaps 133;
QY	4	EEBEE-----EEDGDABETQDSEDEDEM-----EEDDDSDYPREMEDDDDDASYCT- 51
DB	2614	EEKELASAKKQGDQIEQSKQSEVSEVVAEIKISGKIIEPKKPEM-DTEAKSEKATV 2672
QY	52	-----ESSFRSHSYSTSPGRRKPRV-----HRRSP- 78
DB	2673	LDRQVLEKELEASAEKQGDQDQVEKKSQKPEVSEVVAEKISEETIEEPKPEVKDTEIKS 2732
QY	79	-----ILLEKDIPLPEFPKSSDLAVPNE--HIMNVIAIYEVLRNFGTVLRLSPPR 127
DB	2733	EKATALDKQVLEKELEASAKQGDQDQVEKKSQKPEVSEIVA-----EKISEKT 2781
QY	128	FEDFCAALYSQEOCTLMAEMHVLLKAVLREE--DTSNTTFFGADLKDSVNSTLYTFIDGM 185
DB	2782	IEEPKPEVKDTE--IKSEKATALDKQVLEKELEASAKQGDQDQVEKKSQK----- 2831
QY	186	TWPEVLRVYCESKEYHHVLPYQEAEDYPYGPVENKIKVLQFLVQDQFLATNTIAREBLMSE 245
DB	2832	--PEVSEVVAEKISE-----ETIEEPKPEVKETE-----KSEKATVLDK 2870
QY	246	GVIOYDDHCRVCHKGLDCCETCSAVYHLECVKPPLEEVPEDEWQCEV--CVAHKVPGV 303
DB	2871	QVLEKELEASAKQGDQ-----QDVEKKFQKAESEVVAEKISE 2910
QY	304	TDCVATIQNKPIRHEPIGYDRSRKYWFLNRRLLIEEDTE-----NENEKKIWIYSTKV 359
DB	2911	T--IEE--PKKPEVKDTEIKSEKATA---LDKQVLEKELEASAKQGDQDQVEKKSQRP 2962
QY	360	QLAELIDCLDKVWEALCKILEEMREIHRMD--ITEDLTNKAAGNSKSFLLAAANEEL 418
DB	2963	EVSEVI-----AE--KISEKIEEPKPEKETEVEKSEKATVLDKQVL---BEKEL 3008
QY	419	ESIRAKKGDID-----NVKSPETEKKDNETENDSDRAENREE 457
DB	3009	EASAKQGDQDQVEKKSQKPEVSEVVAEKVSEKIEPKPEVKETEAKSE--KATTLDMQV 3067
QY	458	FEDQSLKSDDDKTDPDDPEQ--GKSEVGDFKSEKSNGLSESPGAGKASGASTRIITRL 515
DB	3068	LEERELEA--SAQKQGDQVEKKSQKPEVSEVVAEKISEEKIEP----- 3110
QY	516	RNPDSKLSOLKSOQVAAAHAANKLFKEQEV--LVVNSOG--BISRLSTKKEY----- 565
DB	3111	KKPEKETEYKSEK---ATVLDKQVLEKELEASAKQGDQDQVEKKSQKPEVSEVVAEK 3166
QY	566	IMKGNINNYFKL-----GQEGKYR-----VYH 587
DB	3167	VSEKIEEPKPEVKETEYKSEKATVLDKQVLEKELEASAKQGDQDQKSRDRIITKLK 3226
QY	588	NOYSTNSFALNK-----HQHRE-----DHDK-----RRHLAKFCLTPAGEFKWN 627
DB	3227	ERLTLSKALGSSVDEILRESREIVNNLEDKVVAKHLFKLRDHIYHTYDGRKEENK-- 3284
QY	628	GSVHGSKVLITSLRTITOLENNIPSSFLH-----PNWASH--RANWIKAVOMCSKPKE 680
DB	3285	-----EKELFESPIELCEASPEAAEKVKIYNLKEIK-----TN 3318
QY	681	FALALAILCA-----VKPVMLPIWRBFLGTHRLHRMTSIEREEKEVKKK--EKKQ 731
DB	3319	VILTAKTIQIIDDNSMFTAPSLIP-----KLNLER--VAVKIQSETYVDKSS 3365
QY	732	EEBETMQQATWVKYT-----FPVKHQVWKQGEYRVTYGGWSWISKTHYRVFVKPLPG 786
DB	3366	KEMISLQQSLMDIFVILDDFDDTEVLPKPIENIKTLLSDYDYLEKKDGPLLTAVING 3425
QY	787	NTNV-----NTRKSLGKTNMNMDESKRSRSPKKIKIE--PDSEKDEWKGSDA--- 838
DB	3426	KINWVSOHILTIIEEVK--QLTENHDQKEDKVDVNAEADNFADEKRESQKEEKIDSEAKHK 3484
QY	839	-AKGADONEMDISKITEKKDQVKELLDSDSD----- 869

DB	3485	KSVSEKKSIEEBKLEDDKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKG 3544
QY	870	-----KPCKEEPMVDDMKTESHVNCQESSQVDVYN--VSEGF---HLRTSYK 913
DB	3545	EAKPKKAKVLEKKSIEEBKLEDDKKEKQTESAID--EKSQKAEVSEIVSEKITDEKAQESQK 3603
QY	914	KTKTSKLGDLERRIKQFTLEBKQRLKIKLGGIGKIGKIGTSTNSKNLSESPVITKAK 973
DB	3604	KEYKDS-----EAKPKKAKVLEKKSIEEAKLE-----DKKETQDTSADERSQKAEVS 3651
QY	974	EGCOSDSMRQOSPNANNOQPEDLIOGCSQSDSSVLRMSDPSTHTNKLYPKDRVLDVSI 1033
DB	3652	E-TVSEKITDEKAQESQKEVKD-----SEAKPK---KAKVLEKKS 3689
QY	1034	RSPETCKPKONSTIENDIEEK-----VSDLAS-----RGOEPTKSTKGNDFIDDSKLA 1082
DB	3690	EEBKLEDDKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAVL 3749
QY	1083	SADDITGLICKNKKPLIOEESDTIV--SSSKSALHSVPKSTNDRDATPLSRAMDFEGKL 1140
DB	3750	EKKSIEEBKLEDDK---ETQDTSADIDEKSQKAEVSEIVSEKITD-----EKAQESQKEE 3800
QY	1141	GDSESNSTLENSSDTVSITODSSEEDMIYONSNESISEQPTREODVEVLEPLKCLVSG 1200
DB	3801	VKDSEAKPKKAKVLEKKSIEEBKLEDDKKEKQTESAIDE---KSQKAEVSE--IVSEKITD 3855
QY	1201	ESTGNCNEDRLPVKTEANGKKPSQOKKLEERPV--NKCSDOIKLKNT---TDKKNENRE 1255
DB	3856	EKAQESQK--EYKGSSEA---KPKKAKVLEKKSIEEBKLEDDK--KEKOTESAIDEKSQKAEV 3910
QY	1256	SE-----KKQRTSTFOINGKDNKPKIYLKGECLKEISESRVSVGNVPEKVNINKI 1307
DB	3911	SEIVSEKITDEKAQESQMBEVKDSSEAKPK---KAKVLEKKSIE-----EERLENKKEK 3960
QY	1308	IPENDIKSTVKSARPFINGDVIMEDFNERNSSSETKSHLLSSDAEGNYRDLSE--TL 1365
DB	3961	QTESAIDEKSQKAEV-----SEIVSEKITDEKAQESQKKEVKDSEAKPKKAVLEKKS 4014
QY	1366	PSTKESDSTQTTTPPSACSPESNSV-----NOVEDMEIETSEVKKVTSPITSEESNL 1418
DB	4015	EEBKLEDDKKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKKEVKDSEAKPKKAVL 4074
QY	1419	SNDFIDENGLPINKNENVAGESKRVITVETVMTSTVATESKTVTKVEGDKQTVVSS 1478
DB	4075	EKKSIEEBKLEDDKQDTSADIDEKSQKAEV---SEIVSEKITDEKAQESQKEEYKDS 4130
QY	1479	ENCAKSTV---TTTTTTVTKLSTPSTGGSDVIISVKEOSKTVVTTVTDSLTGTTGTV 1534
DB	4131	AKPKKAKVLEKKSIEEBKLEDDKKEKQTESAIDEKSQKAEVSEIVSEINIDE----- 4181
QY	1535	TSMTVSKYSTRD-----KVKLM-----KFSRPKTRSGTALPSYRKFTKSTKKS 1580
DB	4182	-KAQESQKKEVKDSEAKPKKAKVLEKKSIEEBKLEDDKKEKQTESAIDE---KSQKAE 4234
QY	1581	IFVLPNDLL--KKLARKGGIREVPYFYNNAKAPALDIWPPSPRPTFGITWRYRLQTVKSLA 1639
DB	4235	VSEIVSEITEEKAQESQKKEVK--DSKAKPK-----KAKVLEKKSIE 4275
QY	1640	GVSLMLRLWLASLRWDMAAKVPPGGGSTRTETSETTEITTEILIKRRDVPYGIREFYCI 1699
DB	4276	EAKL-----EDKKEI--QDTSADIDEKSQKAEVS-----EIVS 4305
QY	1700	KIICPIGVPETPKETPTPQRKGLRSSALRPKRPETPKQTPGVPIETWVAEELEWEIR 1759
DB	4306	EKI-----TDEKAQESQKEVKDSEAKPKKAK-----VLEKKSIEEKELE----- 4345
QY	1760	AFAERVEKEKAQAVEQQAQRLEQQRPVIAVSTTSTSTSTISPAQKVMVAPISGSV 1819
DB	4346	---NKKKETESAIDEKSQK-----AEVSEIVSEKITDEKAQESQKKEVKG-- 4389
QY	1820	TTGCKNVLITKVGSPATVTFQONKNPHQTFATVWVKOGSNGVVOQKV--LG1IPSSTG 1878

```
Db 498 VQ--MPDGKHL-----VLTSTSSAGQGNKMKVPIKP----- 529
Qy 2118 PAQSSSVGPAQAQOTAPSAARPQOTOPQSPAQEVOTQEVQQTIVSSHPVSEAOPT 2177
Db 530 --ASTSSSPATSSAQT--TNPVTPVVKIOIAVKHVKTKNSATQSIASSSRVALP 578
Qy 2178 HAQSKPKVAAQSPQSNVQGSQSPVRVQSPSQ-----TRIRPSPSPQSL 2221
Db 579 LAQIRKLLAQQQQQST--SSSPATSSSPVQKIVSKVNVNTSTSGTQLQOVVQSGSKLV 636
Qy 2222 PQQQSOVQTTTSQPIPIQPHTSLSQPSQSPQVQSSQTQLSSGQTLNNOVSVSPSR 2281
Db 637 VQONAAQGGKVIISTSAQAQQGTSPPVQQQLVQSQP-IQOSPQOIS--MTQQQIVVGG--- 690
Qy 2282 POLQIQOOPQPOPIAVPQLOQQOVQLSQTQSQVVAQIQAQSGVPQIQILQPIQIQSSA 2341
Db 691 -ORITLSP-GQPIVTRNVNPOSQALQVQQQIQTQOQQOHHV-----VQPPQOFV 739
Qy 2342 VOTHQIQNVTVQAASVQEQLRVQQLRQDQKKQOQIQEIKREHTLQASNQSELIQKV 2401
Db 740 VQSNQI-----VQ-SSPSAQTKLVQLVVQQ--SQQTIEEKTQITTTDSNETG--TQOV 789
Qy 2402 YMKHNAVIEHLKQKSWTPAEE-----NORMIVCNQVMKYILD----- 2441
Db 790 LVPNSTLAQQLAQQAGQLQVATVNGQQVIVKPLGNNOAQIVAH--IKHOGDGNHIVTNSA 847
Qy 2442 -----KIDKEKQAAKKRKREESVQKR----- 2464
Db 848 TAVPOANPQTSVPVQQAQLPQSPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 907
Qy 2465 -SKQATKLSALLFKHK-----EQLRAEILKK-----RALLDKDLQIE-----VQEE 2505
Db 908 PAQQAQSVESLLQNPQPGTVKCVTAQVLTEHGPRIVLQGLVGNDFTAQQLQLVQTO 967
Qy 2506 LKRDLIKKE-----KDLMLAQATAVAPCPVTPV 2537
Db 968 VRQQLMKQAQESKGLVGLPTKTYILAVQENAVQSPPLTPV 1010

RESULT 12
Q9V7G8 PRELIMINARY; PRT: 7210 AA.
ID Q9V7G8
AC Q9V7G8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CG18255 protein
GN STRN-MICK OR CG8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Benson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike K., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrar C., Ferrera S., Fleischmann W.,
```

```
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulev G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphry L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Poilard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.S., Woodruff D.A., Weinstein G.M., Weissbach J.,
RA Wang Z.-Y., Worley K.C., Wu D., Yang S., Yao Q.A., Zheng L.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF003808; AAF58087.2; -.
DR FlyBase; FBgn0013988; Strn-Mick.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 13.
DR SMART; SM00409; Ig; 21.
DR SMART; SM00408; ICG2; 20.
DR PROSITE; PS50835; IG_LIKE; 20.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 7210 AA; 808965 MW; 6B2E7395C07140D2 CRC64;
```

```
QY 1188 EVLEPLKCLVSGSTGNCEDRLPVKGTANGKKPSQOKKLEBRPVNKCSDQIKLKNYTD 1247
Db 985 -----GYLSAKQAHDERKLEE----- 1001
QY 1248 KKNENRESEKKGQRTSTFOINGDKNPKIYLKGECLKEISERVSGNVPEPKVNNINKI 1307
Db 1002 ----- 1001
QY 1308 IPENDIKSLVTKBSAIRPFINGDVIMEDFNERNSETKSHLSSDAEGNYRDSLETPLPS 1367
Db 1002 ----- 1001
QY 1368 TKESDSQTTPSPASCSPESNVQVEDMEIETSEVKVVTSSPITSEESNLNDFIDENG 1427
Db 1002 ----- 1001
QY 1428 LPINKNENVNGESKRKTIVITEVMTSTVATESKTVIKVEKDKQTWVSSSTENCAKSTVT 1487
Db 1002 -----SGVLGEEK----- 1009
QY 1488 TTTTNTVKLSTPSTGSGVDIISVKEOSKTVVTTTVDLSLTTTGGTLVTSMTVSKEYSTRD 1547
Db 1010 ----- 1009
QY 1548 KVKLMFESRPKTRSGTALPSYRKFTVKSTKKSIFVLPNDDLKKLARKKGIREV--PYFN 1605
Db 1010 -----AMPPEIQTFTSKRGKKSIFVLQKKILROMINGGCGOQVMPGFS 1054
QY 1606 YNAKPALDIWPPSPRPFTGWTWRYLQTVKSLAGVSLMLRLWASLRWDDMAAKVPPGG 1665
Db 1055 AGIKSNLLIWPYPAPRTDLCKWQTLNARSLHVALQLKIWSIKFNEFD---PDDT 1111
QY 1666 GSTRTETSET--EITTEIIKRDDVGPYIRFEYCIRKICPTI-GVPETPKETPTPQRKG 1722
Db 1112 HPDRRVVIDPDSHDERRIIRHKEMPPYQGYERYEMEIEIPLYDEPEEDESWSLRNG 1171
QY 1723 -----LRSSALPKRPET-----PKQT-----GPVLIETWVAEELE 1754
Db 1172 GSSEFSHRSSARKRPQRHEFTLSLFGNYPKSKNAFRSLDNRRATAIRREWVDGVTLK 1231
QY 1755 LWE-----IRAFARVEKEKAQV-----EQAKKRLQKQKPTVIATSTTSPST 1800
Db 1232 VFEIKDYKWIARAEAKTAKKLEATKKAQKAKEDERRIIOQOQSVARIPIV-PHNS- 1289
QY 1801 TSTISPAQKVMVAPISGSV---TTGTMVLTTKVGSPPAVTPOQN-----KNPHQT 1849
Db 1290 ---LIPSERNNV-PYLGSOQORRPNNGERGFLEKYNNSSVSPQAHGYASTPPPGYHQ 1345
QY 1850 ATWVKQGSNSGVVQVQKVLGIIIPSTGTSQQTFTSFQPRATVIRPNTSGSGGTSN 1909
Db 1346 PNTIROAGYNQ-----LPRKPTTSPNFOS-RP-VATIPTPOLRAAAGADG- 1390
QY 1910 SQVITGPQIRPG---MTVIRTLQOSTLGLKAIIRTPVMVOPCAPOQVMTIIRGO--- 1961
Db 1391 --VVRAMVTPGNKSVTNTSTPYQ-----ALNRQYQLORQOQOPAVRLTNGHFMD 1443
QY 1962 -----PVSTAVSAPNTVSSTPGQKSLTSAT----- 1986
Db 1444 GTRMGGRNPSVQMQRLPQNRALQRPFGESTTEMRRTVEAALPDNDGDEQPPVPIRYD 1503
QY 1987 STSNIQSSASQPPRQOQVQKLTMAQLTQTOGHGNGOGLTVVIQGGQTTGOLQL----- 2042
Db 1504 PTSNFDQAQQAQHPQSPPYISTPAQIMRTPOGGVKH--NVILMKASDGTOKRWLKPQ 1561
QY 2043 IPOGVTVLPQPGQOLMQAAMPNGTVQRELTPLATTATTASTTTTTSITTAAGTGQRQS 2102
Db 1562 FPPEGTVI--STGQVVVYRQPTAVQORQLYT-----ATPGTRVIRIPNANGAPRQDDH 1613
QY 2103 KL-----SPO-MQVHODKTLPP-----AQSSSVGPAKAQPOQAQPSARPO---- 2141
Db 1614 QVMRRVVQASGPRAEMVDDQGTTPPGQVRYVLOGGNSGTPNVPKVVSSRCGPRGGLT 1673
```

```
QY 2142 ----PQTQPOSPAQ-----PEVOTQPEVQTOQTTVSSHPVSEAQ 2175
Db 1674 MQMVQOQQQHNPFAHYDMPPDDATGFAVSTTT---EQVPDEQQ 1713

RESULT 11
QY 960Y3 PRELIMINARY; PRT; 1022 AA.
ID Q960Y3 AC Q960Y3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE LD30146P
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051776; AAK93200.1;
DR FlyBase: FBgn0000541; E(bx).
SQ SEQUENCE 1022 AA; 111667 MW; 1174939B61962E63 CRC64;

Query Match 4.7%; Score 674.5; DB 5; Length 1022;
Best Local Similarity 25.5%; Pred. No. 5.1e-21;
Matches 286; Conservative 144; Mismatches 361; Indels 332; Gaps 45;

QY 1535 TSMVTSKEYSTRDKVLMKFSRPKTRSGTAL---PSYRK-----FVTSTKTSFVLP 1585
Db 100 TSSPRGRVYLLNDAKL--YEQAVKTEDKSTITTKPSYSRYPLISNFLTTHKKRSLVLP 157
QY 1586 NDCLKKLARKGGTREVYPFYNNAKPALDIWPPSPRPFTGWTWRYLQTVKSLAGVSLML 1645
Db 158 RFELLKLRLGKSSNGFHHAAKN--TIWQYQCSRLPFTCSYRSTNATSLSLALQL 216
QY 1646 RLWASLRWDDMAKVPVGGGSTRTETSEITEITTEIIKRDDVGPYIRFEYCIRKICP 1705
Db 217 RILWLCRLWDDMLAK-PPSTDGKHQVTTDEIVTEIYLLKLRHSGRYGKTSYLRKRWIP 275
QY 1706 IGVETPKETPTPQKGLRSSALRPKRPETPKOTGPVLIETWVAEELELWEIRAFERV 1765
Db 276 LEMPKTVREV-TSIRSGLR----KKRAESPQTEPQITEEWDEDEKLELWEIKFMGEKQ 330
QY 1766 EKEKAQAVEQ--AKKRLQOKPTVIATSTTSPSTTSITSPAKVMVAP-IGSVTTGT 1823
Db 331 EKARLSAVTRESVARQLE-----ASGNSGNTSINGALGVAGRVQLAKLSELDVK--E 381
QY 1824 KVLVTTKVGSPATVTTQKNKFHOTFATWVKQGSNSGVVQVQKVLG-----IIPSTGT 1879
Db 382 KMEQQLKL---QRAVHQQRK-----LVATGEITRSVTPVRKGVIGSRVIVKPNPDGT 430
QY 1880 SQQFTTFSFQPRATVIRPNTSGSGGTTNSQVITGPIRPGMTVIRTPLOQSTLQKAI 1939
Db 431 TR----- 434
QY 1940 RTPVMVOPGAQPOQVMTIIRGQPVSTAVSAPNTVSSTPGQKSLTSATSTSNIOSSASOPP 1999
Db 435 -----QAQVTVQVSTGGANTAAA---ASPTVG-----GSTSTQS--NPSTSTPH 474
QY 2000 RPOQGVKLTMAQLTQTOGHGNGOGLTVVIQGGQTTGOLQIIPQGVTVLPG--PQQL 2057
Db 475 KVQ-----IIRG-----PDGKSVSRGLNPGQQL 497
QY 2058 MQAAMPNGTVQRFVLTPLATTATTASTTTTSTTAAGTGQRQSKLSQPMQVHQDKTLP 2117
```



```
Db 4 GPAAQPPQAPAAQPP--QPQPAQPEVQTOP-----AVSSHPSETQPSQAQTSKP 55
QY 2185 QVAAQSPQSNVQSGSPVRVQSPQTRIRPSTPSQLSPGQSQVQTTSQPIPIQHTSL 2244
Db 56 LVATQCPQSSVQSGSPVRVQSPPLIRICPSTPSQVTPGQPPQVQTASQPIPIPTSL 115
QY 2245 QIPSGQSPQSPQVQSTQTLSSGQTLNVSVPSPRPQIQPQVIAVPOLOQVQ 2304
Db 116 QAFSGQSPQSPQVQSTQTLSSGQTLNQVTVLSPSCP-----QPQPVIAVQOL-QVQ 169
QY 2305 VLSQISQVVAQIAQAOQSGVPQIKQLPIQIQOQSSAVQTHQIQNVVTVQAAQVQOLQ 2364
Db 170 VLSQISQVVAQIAQAOQSGVPQIKQLPIQIQOQSSAAQT---QSVTVQAAQVQOLQ 226
QY 2365 VQOLRQOQKQKQOIEIKREHTLQASNOSEIIQOVVMKHNVAIEHLKOKKMTPAERE 2424
Db 227 VQOLRQOQKQK--QOIEETERHTLQASNOSEIIQOVVMKHNVAIEHLKOKKMTPAERE 285
QY 2425 ENQRMIVCQVMKYIILDKIDKEEKQAAKKRREESVEQKRKQNAATKLSALLFKHKEQLR 2484
Db 286 ENQRMIVCQVMKYIILDKIDKEEKQAAKKRREESVEQKRKQNAATKLSALLFKHKEQLK 345
QY 2485 AEILKRALDKDLQIEVQBELKRLKIKKEDLMQIAQATAVAAPCPVTPVLPAP-PA 2543
Db 346 AEILKRALDKDLQIEVQBELKRLKIKKEDLMQIAQATAVAAPCPVTPVLPAPAPA 405
QY 2544 PPSPPP--PPGVQHTGLLSTPT--LPVASQKRREEEKSSSKKKKMTSTSTETKK 2599
Db 406 APPAPRPPSTHSLPPAGHPAPLPVTSQKRREEEKD--SKSKKKKMTSTSTETKK 463
QY 2600 DTRLYCICKTPYDESKFYIGCDRCQNWYHGRVGLQSEAEILDEYVCPQCSQSTEDAMTV 2659
Db 464 DTRLYCICKTPYDESKFYIGCDRCQNWYHGRVGLQSEADLIDEYVCPQCSQSTEDAMTV 523
QY 2660 LTPLETEKDEGLKRLVRSLOAHKMAWPFLEVPDNDAPDYGVYIKPMDLATMEERVQR 2719
Db 524 LTPLETEKDEGLKRLVRSLOAHKMAWPFLEVPDNDAPDYGVYIKPMDLATMEERIQR 583
QY 2720 YYEKLTFEVDATKIFDNCRYNPNPSDPFYQCAEVLESFVQKLGFKASRSHNNKLOST 2779
Db 584 YYEKLTFEVDATKIFDNCRYNPNRTPFYQCAEVLESFVQKLGFKASRSHNNKLOST 643
QY 2780 A 2780
Db 644 A 644

RESULT 8
Q9H5E0
ID Q9H5E0 PRELIMINARY; PRT; 412 AA.
AC Q9H5E0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ23531.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA Oiyasahi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RL "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027184; BAB15686.1;
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 46539 MW; D72A6D830BB12B94 CRC64;
```

```
Query Match 9.5%; Score 1368.5; DB 4; Length 412;
Best Local Similarity 57.0%; Pred. No. 3e-51;
Matches 302; Conservative 18; Mismatches 41; Indels 169; Gaps 7;
QY 2107 QMOVHOD-----KTLPPAQSSSVGPAKAQPO-----TAQPSARPQPTQPOSP 2149
Db 31 QOVNKGVRNCHPCRYIKTKPCHQLSHQVWVQOKPSHRLLSQLSPSPKPSPOLSLK 90
QY 2150 AQPEVQTOP--EVQQTQTVSSHVSEAPQTHAQSSKQVAAQSPQSNVQSGSPVRVQSPS 2208
Db 91 FRLSLFRKQLFPHMSLLKHNP---PHTSHFS-PKLOSQSPQSNVQSGSPVRVQSPS 145
QY 2209 QTRIRPSTPSQLSPGQSQVQTTSQPIPIQHTSLQIPSGQSPQVQSTQTLSSG 2268
Db 146 QTRIRPSTPSQLSPGQSQVQTTSQPIPIQHTSLQIPSGQSPQVQSTQTLSSG 193
QY 2269 QTLNVQSVSPSPRQIQIQPQVIAVPOLOQVQVLSQISQVVAQIAQAOQSGVPQOI 2328
Db 194 ----- 193
QY 2329 KLQLPQIQOQSSAVQTHQIQNVVTVQAAQVQEQQLRVQQLDQOKKKQOQIEIKREHTL 2388
Db 194 ----- 193
QY 2389 QASNQSEIIQOVVMKHNVAIEHLKOKKMSMTPAEREENQRMIVCQVMKYIILDKIDKEEK 2448
Db 194 -----QVVMKHNVAIEHLKOKKMSMTPAEREENQRMIVCQVMKYIILDKIDKEEK 242
QY 2449 QAAKKRREESVEQKRKQNAATKLSALLFKHKEQLRAEILKRALDKDLQIEVQBELKR 2508
Db 243 QAAKKRREESVEQKRKQNAATKLSALLFKHKEQLRAEILKRALDKDLQIEVQBELKR 302
QY 2509 DLKIKKEKDLMLQIAQATAVAAPCPVTP---PVLPAAPPAPPPPPPGVQHTGLLSTPTL 2565
Db 303 DLKIKKEKDLMLQIAQATAVAAPCPVTPPAPPAAPPAPPPPPPPPAVQHTGLLSTPTL 362
QY 2566 PVASQKRREEEKSSSKKKKMTSTSTETKKDKLYCICKTPYDESK 2615
Db 363 PAASQKRREEEKSSSKKKKMTSTSTETKKDKLYCICKTPYDESK 412

RESULT 9
Q45409
ID Q45409 PRELIMINARY; PRT; 1711 AA.
AC Q45409;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE F26H11.2a protein.
GN F26H11.2 OR F26H11.2A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC Barlow K.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81515; CAB04197.1;
DR WormPep; F26H11.2a; CEI5908.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02791; DDT; 1.
DR SMART; PF00628; PHD; 1.
DR SMART; SM00384; AT_hook; 1.
```



```
DE Nucleosome remodeling factor large subunit NURF301.
GN E(BX) OR NURF301 OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21468388; PubMed=11583616;
RA Xiao H., Sandaitzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,
RA Fu D., Wu C.;
RT "Dual functions of largest nurf subunit nurf301 in nucleosome sliding
RT and transcription factor interactions.";
RL Mol. Cell 8:531-543(2001).
DR EMBL; AF417921; AAL1664.1; -.
DR FlyBase; FBgn000541; E(bx).
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR00345; CytC_heme_bind.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 3.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2669 AA; 300687 MW; 6B4925AFF489D6F CRC64;

Query Match 19.8%; Score 2839.5; DB 5; Length 2669;
Best Local Similarity 27.5%; Pred. No. 2.6e-113;
Matches 869; Conservative 383; Mismatches 919; Indels 985; Gaps 98;

QY 4 EEEEDGGAETQDSEDEDEDEDEDDSDYPPEMEDDDDDASCTESSFRSHSTYSS 63
DB ESEYHYGDFGDEEDKSDNEDDMLLTPSDDESLEVANESEFSVC---SFQNGV--- 152
QY 64 TPCRRPRVHRPSPILEE-KDIPPLEFPKSSDLMPVNEHINVAIYEVLFNFGTVLR 122
DB ---GRPPRPSPVPMVQGRQVAALDLPDSSDDLFIANTHVLRALSIYEVLFNFRHWR 209
QY 123 LSPFREFCAALVSQEQCTLMAEMHVLLKAYLREEDTSNTTFGPADLKDSYNTIYFI 182
DB LSPFREFDCAALACEQSALLTEVHIMLLKALIREDAQGTGFGPLDQDVTNISLYLI 269
QY 183 DGMTWPEVLRVYCESKEY-----HHVLPQEAEDYPYGPVENKIKVLQFLVDQFLTNTIA 238
DB LSPFREFDCAALACEQSALLTEVHIMLLKALIREDAQGTGFGPLDQDVTNISLYLI 269
QY 270 DSITWPEVLRVYVESDKTFDRNFVHLSHTE---YPYTYGYNLEVLQFLSDQFLTSNSI 326
QY 239 REELMSEGVIQYDDHCRVCHKGLDCCCTCSAVYHLECVKPPLEVPEDWOCCEVVAH 298
DB REELMSEGVIQYDDHCRVCHKGLDCCCTCPAVYHLECVKPPLEVPEDWOCCEVVAH 298
QY 327 RDVLMGEPHYDDHCRVCHKGLDCCCTCPAVYHLECVKPPLEVPEDWOCCEVVAH 386
QY 299 KVPGVTDCAEIOKKNPIYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKIKWYSTK 358
DB KVPGVTDCAEIOKKNPIYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKIKWYSTK 358
QY 387 KVSQVDCVLPQEQGVLRHDSIGVDHRGKRYWFIARRIFI-EDQENFT---CWYSTT 442
QY 359 VQAEILDCIDQWYAECLKILEEMREEIHRIMDITEDLTNKAAGSNKSFLLAAANEIL 418
DB VQAEILDCIDQWYAECLKILEEMREEIHRIMDITEDLTNKAAGSNKSFLLAAANEIL 418
QY 443 SKLKLILLSLDAELETFLHSQITERRDETERQMKLTETLTNKHKHTKRSVI----- 494
QY 419 ESIRAKKGGIDNVKSPETEKNETENDSKDAEKNREEDQSLKSDDKTPDDDPEQ 478
DB ESIRAKKGGIDNVKSPETEKNETENDSKDAEKNREEDQSLKSDDKTPDDDPEQ 478
QY 495 -----ETEQEAKN-----ELLEKEVLEDEK----- 516
QY 479 GKSEVGDFKSEKNGELSEPGAGKSGSTRIITRLRNPDLSQLKSOQVAAAHEAN 538
```

```
DB 517 -----GDAKSESQIE-----GKKQBECKMVR-----QKSNQLT----- 547
QY 539 KLFKEGKGVLVVNSQGEISRLSTKKKEVINMKNINNVFKLQGBQEKYRVYHNOYQSTNSFALN 598
DB 548 -----NGLTFKLGMQEGFKNVYNOYSTNPFIALN 576
QY 599 KQHREDHDKRRHLAHKFCLTPAGEFKWNGSVHGSKVLTIISTLRLITOLENNIPSSFLH 658
DB 577 KQORNEERDKRRHLSHKFSLT'TASDFKWGITMGTTDNMTITLRLQFLINFSNIAASFLN 636
QY 659 PNWASHRANWIKAVOMCSKPREFALALALECAKVPVMLPIWREFLGRHLRMTSIER 718
DB 637 INWVNKKIWNAAVMNARRPSEFAVLLLPQASLKSVMZANVHVEQLGHTTLQRTISAER 696
QY 719 EEKEKVKKKEKKQ---EEETHMOQATWVYTPFKVQVVKQKGEERYVYGGWSWISKT 775
DB 697 EERKKLEKREKREDEEERNRLAFNYIKYTLGLKHQVWKQKGEERYVHVGOWGLWSS 756
QY 776 -----HVYRFVPLPGNTNVNRYKSLLEGTKNNHNDENDESDDRKRKSRPKIKIEPSEK 830
DB 757 RRCGVARRAQPLTHNRVYVHYTM---GEENDVNEII-----LVDPTQREFMQOSES 805
QY 831 DEVKGSDAAKGADQ---NEMDISKITEKDDQVYKELLDSDSDKPCKEPMEVDDDMKTESH 888
DB 806 SNVDGQVCHYLPDQYKNVKVIEDVTE-----KTKGH 836
QY 889 VNCQESSQVDVNVNVEGPHL--RTSYKKTKSKLGLLERRIKQFTLEKORLEKIKLE 946
DB 837 -----IDVSKALNAPGRTYYSKVARSLDDLDLRUKLAEVE--QMASKI--- 881
QY 947 GGIKGIGKTSTNSSKNLSESPVITKAQEGCQSDMRQEOPNANNQDPEDLIQCSQSDS 1006
DB 882 -----PSDM 885
QY 1007 SVLRMSDPSTTNKLYPKORVLDVVSIRSPETKCPKONSTENDIEEKVSDLASRGEPTK 1066
DB 886 KPLLVSQNNNTANS-----KQTFLE-----KRLRLTE 913
QY 1067 SKTKGNDFFIDDDSKLASADDIGTICKNKKPLQIEESDTIVSSKSALHSSVPKSTNDRD 1126
DB 914 VQAKG----- 918
QY 1127 ATPLSRAMDFEGKGLCDSESNSTLSSDTVSQDSEEDMIVQNSNESISEQFTREOD 1186
DB 919 -----GPANVLELVNSLAKIQIVRLQES-----QL 945
QY 1187 VEVLEPLKCELVSGESTGNCEDRLPVKGTANGKPKSQOKKLEERPYNKSDQIKLKNIT 1246
DB 946 NREAKVFR-----YTKEC-----NTNSNAVSIQI-----NTCYSPLCLOKAR 984
QY 1247 DKN-----NENRESEKKGORTSTFQINGKDNKPKIYLGECLEISESRVSGNVPEKVN 1302
DB 985 AKKELLKLLKAHTAGNGSKETVAAILGAVKKFSIL-----EOKITEGRE--- 1030
QY 1303 NINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETSKHLLSSSDAEG----- 1356
DB 1031 -----STQVAVDDSEEGKPAESAPLDLLQDWEHARAHAV 1065
QY 1357 NYRDSL-----ETLPSTK-----ESDSTQTTTPSACPSNSNVQEDM-----E 1396
DB 1066 PFSDSLITECILVDQECVTNTKIQEYNASSGCGNTTPDSNTQSDKIDYTESMDVCSNVE 1125
QY 1397 IETSEVKVTSPITSEESNL-----SNDFID---ENGLPINKNEN 1435
DB 1126 IESTESIVTGLSGNAEDVDMPGWRKRNQKSKSYIGTKVDQDQTLDRKIDPLNK--- 1182
QY 1436 VNGESKRKTVITEVTMTSTVATESKTVIKVEKGDQKQTVVSSSTENCAKSTVTTTTVTVK 1495
DB 1183 -----QNRFPFIT-----ARPVKREC-----VKYERETFEENGNERV----- 1214
QY 1496 LSPSTGGSV-----DIISVKEQSKTVVTTVTDSLTTGTLTMTVSKYSTKSTRDKVKL 1551
DB 1183 -----VKYERETFEENGNERV----- 1214
```

Db 914 VQAKG----- 918
 QY 1127 ATPLSRAMDEGKLGCDSESNSTLENSDSTVSTODSEEDMIIVONSNESISEOFRTREOD 1186
 Db 919 -----GPANVNLNLVSLAKIQIVRLQFS-----QL 945
 QY 1187 VEVLPLKCLVSGESGNCEDRLPVKGTETANGKPKSQOKKLEBRPVNKCSDOIKLKNTT 1246
 Db 946 NREKAFVRC-----YKCE-----NTNSNAVSOITQ-----NTCYSPLCLQKAR 984
 QY 1247 DKKN-----NENRESEKKGORTSTFQINGKDNKPKIYLKGBCLKEISRSRVVSGNVEPKVN 1302
 Db 985 AKRELLLLLKAHTAGNSKETVAAILGAVKKPSIL-----EQLTEGKRE----- 1030
 QY 1303 NINKIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNNSSETKSHLLSSSDAEG----- 1356
 Db 1031 -----STQVAVDDSEEGKPAESEAPLDLLQDWEHARAHAV 1065
 QY 1357 NYRDSL-----ETLPSTK-----ESDSTOTTPPSASCPSNSVNOVEDM-----E 1396
 Db 1066 PFSDSLITELIVDQECVTNKKIQEVNASSGCTTPDSNTQDSDKIDYIESMDVCSNVE 1125
 QY 1397 IETSEVKKVTSPITSEESNL-----SNDFID-----ENGLPINKNEN 1435
 Db 1126 IESTEDSIVTGLNSGNAEDVMDPGWRKRKNQSKSYICTKDVLDOTLDDKDIPLNK--- 1182
 QY 1436 VNGESKRKVTITEVMTSTVATESKTVIKVEKGDQTVVYSTENCACAKSTVTTTTVTVK 1495
 Db 1183 ---QNRREPIT---ARVPKREC---VKYERETFENGNERV----- 1214
 QY 1496 LSTPSTGGSV---DIISVKEOSKTVVTVTTVDSLTITGTLVTSMTVKEYSTRDKVKL 1551
 Db 1215 YTSRPRGRVYLNDAAKLYEQA-----VKTEDKSTI 1246
 QY 1552 MKFSRPRKTSRGALPSYRFVTKSKTSIFVLNDDKLARKGGIREVPYFNYNAKPA 1611
 Db 1247 TK--KPSYSR---YPLISNFLTHKKRSLVLPREFELLKLARLGKSGSTNGFHHAAKN 1300
 QY 1612 LDITWPPSPRPTGIVWRYRLQVKSLAGVSLMLRLWASLRWDDMAAKVPPGGGSTRTE 1671
 Db 1301 -TIWOYQCSPLPRTCWSYTSNATSLSSIALQRLWLSCLRWDDMIAT--PPSTDGKHQV 1358
 QY 1672 TSETETTTTIIKRROVGPVIRFEXCIRKIIICPIGVPEPTEPTTPORGLASSALRPK 1731
 Db 1359 TTDEIVTLELLKLSRGYGETSYLRRKVVIPLEMPKTVREV--TSIRSLGR---KPK 1413
 QY 1732 RPETKQTPGVIIETWAELELWEIRAPAEVEKEKAQAVEQQ--AKKLEQOKPTVIA 1790
 Db 1414 RAESPQTEPQITEEWVDEKLELWEIKFMGEKQKARLSAVTRSVASROLE-----A 1466
 QY 1791 TSTTSPTSTSTISPAQKVMVAP--ISGSVTTGTKMVLTFTKVGSPATVTFQOKNHFQTF 1849
 Db 1467 SSGNSNTSTNGALGVAGRVQLAPKLSDEVK--EKMEQQLKL---ORAVHQKQK----- 1515
 QY 1850 ATWVKOGQNSGVVVOQKVLG---IIPSTGTS---QOTFTSFQ-----PRATVITRP 1998
 Db 1516 --LVATGEITRSTVPKGVIGSRVIRVKNPDGTTIRIIQOAVTVQVSGTGANTAAAAASP 1573
 QY 1899 NTSGSGTTSNS-----QVITGPO-----IRPGMTVIRTP-----LQOSTLGKA 1937
 Db 1574 TVGSGTSTQSNPSTSPHKVQIIRPGDKVSVRGLNPGQQLVQMPDGKLVLTITSSNS 1633
 QY 1938 IITPVMVOPAQPOQVMTQIIRGQPVSTAVSAPNTVSTSTPGQKSLTSATSTSNQSSAQ 1997
 Db 1634 AGQGNKMKVPKIPASTSS---SPAISSAQTTTNPVTPVVIQIAVKHVTKNSATQSIAS- 1688
 QY 1998 PPRPQOGVKLTMAOLTOLTQGHGNGQLTVVLOGOGTQO-----LQILPOGVTV 2049
 Db 1689 -----SSRVALPLAQI-----KNLLLAQOQOQSTSSSPATSSSPVQKRVSKVYN 1733
 QY 2050 LPPGGQQLMAAMPNGTVQRFLETPLATATTATTTTTVSTTAA---GFGEQKQLSP 2106
 Db 1734 TSTSGGTLOQVQVQSGS-----KLTVGQNAQOGKVIISTSAOQOQSTSPVQOQLVQ 1785

QY 2107 OMQVHQDKTILPPAQSS--SYGPAKAQ-----POTAOPSA 2138
 Db 1786 SQPTQOS-----POQISMSTQVGNOPTKVKIOQIVNTSNVQOOIVVGGORIIILSPGOTIVTQ 1841
 QY 2139 RPQOTQOPQSPAQPEVOTOP-----EVQTOTVSSHVPSEAPTHAQSSKQVQAAS 2190
 Db 1842 RNVPQSALQMWQOIQTQOQOQHVVVQOQPFVQSNQIVQSPSAQTKLVKQLVQVQ 1901
 QY 2191 QPQSNVQOGSPVRVQSPSQT-----RIRPSTPSQLSPQOQSVQOTTSQPIPIPHSTSLQ 2245
 Db 1902 QSQOTIEEKQTITTDNETGTQVLPVNSTLAQOLAQGLQVATVNGQVIVKPLGNQ 1961
 QY 2246 ---IPSOQO-----POSOPQVOSSTQTLSSGOTLNOVSVSSPSRQOLQ 2285
 Db 1962 AQIVAHIKHOGDGNHIVTNSATAVPQANPOTSPVRQOALPPQSPQOQVVVQ-----QOQ 2016
 QY 2286 IQQPOP-----QVIAPQLOQOQOV--LSQIOSQ-----VVAQIOA 2319
 Db 2017 IHQOSPNTFESGVTPITQOPVLQVAVQAPAOQOALSVEESLQJNQPPGTVIKCVTAQVLQ 2076
 QY 2320 QQSG-----VPOQIKILQP-----IQIOQSSA 2341
 Db 2077 TEHGPRIVLOGLVGNDFTAQOLQLOVQVKQOLMKAQESNGKLGVLGPTKTYLAVQENA 2136
 QY 2342 VQT-----HQIONVTVQA---ASVOEQLORVQOQLR-----DQOQKKKQOQIE 2381
 Db 2137 VQSQPPPLTPVHQSAHAHQOTNNIEIDATLTATTYEANSTIKDIAINNGDDQENSKCAETE 2196
 QY 2382 IKREHT-----LOASNOSE-----ILOQOV--V 2402
 Db 2197 NSNITTNESFAGTSSLLGSEHDEPTNLAGLIDSETDLENKQNESFVVTVRGYIOKSISNA 2256
 QY 2403 MKHNAVIEHLKQK-----KSMTPABREENORMIV----- 2431
 Db 2257 LKQGNLSPELEEKLVCQKQOENANSTNEWETCSRGVNEALTPSRQTDDETKWIRTSL 2316
 QY 2432 -----CNOVMKYILDKIDKEKQAAKKRKRREESVEQKRKSKONATKLSALLPKHK 2480
 Db 2317 RRPNAMTTSQFNRILK-----KNRSKNDEVAELGEQKOSQ-----LERHK 2357
 QY 2481 EQLRAETILKRALLDKLOJEVOELKRDILKKEKDMOLAOATAVAAPCPVTPVPLA 2540
 Db 2358 ELAKNLTNRKSLRLRLQSEIHE-----DVKTVQVRHVRPLSNA----- 2397
 QY 2541 PPAPPPSPPPPGVQHTGLLSTPLPV-----ASQKRKEEK----- 2578
 Db 2398 -----SPDQSENERSG---EPNLDPKRTVQVQNRHAGRPKKLTKRKEKLYCICRTPY 2448
 QY 2579 -----DSSSKSKKKMMISTTSETKTKOTKLYCICKTPYDES 2614
 Db 2449 DDTKFFYVGCDCSNWFHGDVCSITEEASK--KLSEFICIDCKRARETOOLYCSCRPYDES 2507
 QY 2615 KEYIGCDRCNWHYHGRVCGILOSEAELTIDVYVPCQOSTEDAMTV--LTPLTEKDYEGKLR 2673
 Db 2508 QFTICDCKODWHGRVCGILOSEAEFIDVYVPCQCKKNDANANMKKLTNSNDVEELKN 2567
 QY 2674 VLRSLOAHKNAWPFLEPVDNDPADDYGVYKKEPMDLATMEERVQRRYIEKLTFEFVADMTK 2733
 Db 2568 LIQOMQLHKSAPFMEPVPDKEAPDYKYKKEPMDLKRMEIKLESNTYTKLSEFIGDMTK 2627
 QY 2734 IFDNCRYNPSDPPFYQCAEVLESFFVQKLGKPK 2767
 Db 2628 IFDNCRYNPKESSFYKCAEALSFYVQKIKNFR 2661

RESULT 6

QY5VB8 ID QY5VB8 PRELIMINARY; PRT; 2669 AA.
 AC QY5VB8;
 DT 01-DEC-2001 (TREMREL. 19, Created)
 DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
 DT 01-MAR-2003 (TREMREL. 23, Last annotation update)

QY 1612 LDTWPPSPPTGTTWRYRLQTVKSLAGVSLMLRLILNASLRLWDDMAKVPPGGGSTRTE 1671
Db 1301 -TIWQOCSPLPRTCWSTYTSNATSLSSIALQRLILWSCLRWDWIAK-PPSTDGKHQV 1358
QY 1672 TSETETTTTIIKRRDVGPIRECYCIRKICGIPETPKETPPQKGLRSSALRPK 1731
Db 1359 TTUTEIVTLELLKLRSRGYGETSYLRRKVPIPLEMPKTVREV-TSIRSLGR----KRK 1413
QY 1732 RPETPKQTGVIIETWABEELEWIRAFAERKEKAQAVEQO-AKKRLEQOKPTVIA 1790
Db 1414 RAESPQTEPOITEEWDEKLEWEIKFEMGEQEKARLSAVTRSVASRQLE-----A 1466
QY 1791 TSTSTPSSTTSTISPAQVWAP-LSGSVTTGTCKWVLTTKVSGPATVTFQONKNPHQTF 1949
Db 1467 SGSNGSNTSGALGVAGRVQLAPKLSSEYK--EKMEQQLKL---QRAVHQQRK----- 1515
QY 1850 ATWVKQSGNSGVVQOQVLG---IIPSTGTS---OOTFTSFQ---PRTATVTIRP 1998
Db 1516 --LVATGEITRSTVPKGVIGRSRVIVNPDGTTTRIIOQAVTQVSGTGANATAAASP 1573
QY 1899 NTSGSGGTTNS-----QVITGPQ-----IRPGMTVIRTP-----LQOSTLGRA 1937
Db 1574 TVGSGTSTQSNPSTSPHKVQIIRGPDGKVSVRGLNPGQOLVQMPDGKLVLTITTTSSNS 1633
QY 1938 IIRTPVWQPGAPQOQVMTQIIRQPVSTAVSANTVSSPTPGQKSLTSATSTSNISQASQ 1997
Db 1634 AGQGNKMKVIPKASTSS-----SPAISSAQTTTNPVTPVIKQIAVRHVTKNSATQSTAS- 1688
QY 1998 PPRPOGOVKLTNAQLTQTHGGNGLTVVLOGOGOTTGQ-----LQILPOGVTV 2049
Db 1689 -----SSRVALPQAI-----KNLLILAQQOQOSTSSSPATSSSPVKIVSKVNV 1733
QY 2050 LPPGQOQLMAAMPNGTVORFLEPTLATTATTATTTTSTTAA---GTGEQRQSKLSP 2106
Db 1734 TSTSGOTLQVVFQSGS-----KLIVGQNAQGVKIISTSAAQOQSTSPVQOQOLVQ 1785
QY 2107 QMOVHQDKTLPPAQSS-----VGPAK--AQPTAQPASRPOPTQPOSAPQPEVOTQP 2158
Db 1786 SQPIQGS---POQISMTQOQIIVGGORILSPGIVTQNNVPOSCALQMWQOQIQTOQ 1841
QY 2159 -----EVOTOTTSSHVPSAOPHTHAOSSKPOVAQAQOSQPNQSGOSPVRVQSPSQ 2210
Db 1842 QOQHHVPOQOQVFOVQOQIVQSSPSAOTKLKVLQVQOQOQSTIEEKTQITTTDSNET 1901
QY 2211 -----RIRPSTPSQLSPGQOQVQTTTSQPIQPHITSIQ-----IPSQGQ----- 2251
Db 1902 GTQOVLVPNSTLAQLAQGLQVATVNGQOQIVKPLGNQAOIQAHIKHOGDGNHIVTS 1961
QY 2252 -----POSQPVQOSTQTLSSQTLNOVSVSPSRPQIQIQOQPP----- 2291
Db 1962 NSATAVPOANPOTSPVKQALPPSQPVVQO-----QQIHOQSPTNFESGVTPIITOQ 2016
QY 2292 ---QVIATVPOLOQOVV-LSQIQSO-----VVAQIQAOQSG----- 2323
Db 2017 VLTQVQAOQAQOQALSVEESLLQNPQPGVIVKVCVTAQVILQTEHPRIVLQGLVGNDFTAQ 2076
QY 2324 ---VPQIQKLQPL-----IQIQOQSAVOP-----HQIQ 2348
Db 2077 QLOLVQTVQKQQLMKAQESKGLVLGPTKIYLAQVQENAVQSPPLTPVHQSAHQQT 2136
QY 2349 NVVTVOA---ASVOEQLOQRVQOLR---DOQKKKKQOQIIEKREHT-----LQAS 2391
Db 2137 NNEIDADTLATTYEANSIKDIAINNGDDQNSKCAETENSNTTIFESFAGTSSLLEGS 2196
QY 2392 NOSE-----IIOKOV---VMKNVAVIEHLKQK----- 2415
Db 2197 EHDEPTNLAGLOISETDLENKQNESVFTVRYIQKISNALKQNLSPLEELKVCMQQ 2256
QY 2416 -----KSMTPAERENORMIV-----CNOVMKYIL 2440
Db 2257 QENANSTNEWETCSRVNEALTPSRQTDDETKTISLRPNAMTTSQFNRLIK--- 2313

QY 2441 DKIDKEKQAKKRRKEESVEQKRSKONATKLSALLFKHKEQLRAEILKRALLDKDLQI 2500
Db 2314 -----KNRSKNDVEAEIQQSQ-----LERHKELLKNILKRSLIERNLQS 2357
QY 2501 EVOEELKRLDKIKKEDLMQLAQATAVAAPCPVTPVLPAPPAPPPSPPPGQVHTGLL 2560
Db 2358 EIHE---DVKTVQVRHVRPLSA-----SPDQSENERSG-- 2389
QY 2561 STPTLPV-----ASQKRKEEEK----- 2578
Db 2390 -EPNLDKFRTEVQNPVRHAGRPKKLTKRKEKLYICRTPYDDTKFYVGCDCSNWFHDC 2448
QY 2579 ----DSSSKSKKKMISTTSKTKDKTKLYCICTPYDESKFYICGDRCONWTHGRCVGI 2634
Db 2449 VSITEEASK-KLSEFTICIDCKRARETQQLYCSQCRPYDESOFTYICDCKODWFHRCVGI 2507
QY 2635 LQSEAEILIDRYVPCQOSTEDAMTV-LTPLTEKDYEGLRVLSLOAHKMAWPFLEPVP 2693
Db 2508 LQSEAEFIDEYVCEQCRKNDANAANKKLTNSDVELKNLIKQMLHKSANPFMEPVP 2567
QY 2694 NDADPYGVYKPEMDLATMEERVORRYEKLTEFFVADMTKIFDNCRYNPSDPFFYQCAE 2753
Db 2568 KEAPDYKVIKPEMDLKRMEIKLESNTYTKLSEFIGDMTKIFDNCRYNPKSFYKCAE 2627
QY 2754 VLESFFVQKLKGFK 2767
Db 2628 ALESYEVQIKNFR 2641
RESULT 5
Q9W0T1 ID Q9W0T1 PRELIMINARY; PRT; 2669 AA.
AC Q9W0T1;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE CG32478 protein.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscophila;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA Fouts R.C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K., Nixon K., Nusskern D.R., Paclet J.M.,

Db 61 CATAAGATGCCTGGCCTTCTTGAACCAAGTAGACCCCTAATGATGCACAGATTATTAT 120
Qy 8737 GGTGTTATTAAGAACCTATGGACCTTGGCCACCATGGAAGAAAGATACAAAGACGATAT 8796
Db 121 GGTGTTATTAAGAACCTATGGACCTTGGCCACCATGGAAGAAAGATACAAAGACGATAT 180
Qy 8797 TATGAAGCTGACGGAATTTGGCAGATATGACCAAAATTTTGTATACTGCTGTAC 8856
Db 181 TATGAAGCTGACGGAATTTGGCAGATATGACCAAAATTTTGTATACTGCTGTAC 240
Qy 8857 TACAATCCAAGTGAATTTTACCAGTGTGCAAGCTTCTCGAATCATTTCTTTGTA 8916
Db 241 TACAATCCAAGTGAATTTTACCAGTGTGCAAGCTTCTCGAATCATTTCTTTGTA 300
Qy 8917 CAGAAATGAAGGCTTCAAGTAGCAGGT 8947
Db 301 CAGAAATGAAGGCTTCAAGTAGCAGAT 331

RESULT 10

US-09-920-300A-1410/c
; Sequence 1410, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1410
; LENGTH: 341
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1410

Query Match 3.2%; Score 316.4; DB 10; Length 341;
Best Local Similarity 99.1%; Pred. No. 4.6e-62;
Matches 339; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

Qy 7531 CTGCAACAACAGTCCAGGTTCTCTCAGATCCAGTCCAGGTTGGCTCAGATACAG 7590
Db 341 CTGCAACAACAGTCCAGGTTCTCTCAGATCCAGTCCAGGTTGGCTCAGATACAG 282
Qy 7591 GCTCAGCAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATCCAAATTCAGCAA 7650
Db 281 GCTCAGCAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATCCAAATTCAGCAA 222
Qy 7651 AGCAGTGTCTGCAGACTCCAGGTTTCTCTCAGATCCAGTCCAGGTTGGCTCAGATACAG 7710
Db 221 AGCAGTGTCTGCAGACTCCAGGTTTCTCTCAGATCCAGTCCAGGTTGGCTCAGATACAG 162
Qy 7711 CAAGACAGTTGCAAGGGTTTCCAACTCAGGATCAGGATCAGGATCAGGATCAGGATCAGGAT 7770
Db 161 CAAGACAGTTGCAAGGGTTTCCAACTCAGGATCAGGATCAGGATCAGGATCAGGATCAGGAT 102
Qy 7771 CAGATAGAAATTA-ACGTGAAACACACCTCCAACTTCTAATCAAAAGTTGAAATCATTTCA 7829
Db 101 CAGATAGAAATTAAGGCGTGAACACACCTCCAACTTCTAATCAAAAG-TGAAATCATTTCA 43
Qy 7830 GAAACAGTGGTGTATGAAGCAATAATGCTGTAATAGAACATTT 7871
Db 42 GAAACAGTGGTGTATGAAGCAATAATGCTGTAATAGAACATTT 1

RESULT 11

US-10-099-926-1410/c
; Sequence 1410, Application US/10099926

Publication No. US20030166064A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; APPLICANT: Jjiang, Yuqiu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C2
; CURRENT APPLICATION NUMBER: US/10/099,926
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 1982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1410
; LENGTH: 341
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-099-926-1410

Query Match 3.2%; Score 316.4; DB 12; Length 341;
Best Local Similarity 99.1%; Pred. No. 4.6e-62;
Matches 339; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
Qy 7531 CTGCAACAACAGTCCAGGTTCTCTCAGATCCAGTCCAGGTTGGCTCAGATACAG 7590
Db 341 CTGCAACAACAGTCCAGGTTCTCTCAGATCCAGTCCAGGTTGGCTCAGATACAG 282
Qy 7591 GCTCAGCAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATCCAAATTCAGCAA 7650
Db 281 GCTCAGCAAGTGGTGTGCCCCAGCAAAATCAAACTCCAGTTACCTATCCAAATTCAGCAA 222
Qy 7651 AGCAGTGTCTGCAGACTCCAGGTTTCTCTCAGATCCAGTCCAGGTTGGCTCAGATACAG 7710
Db 221 AGCAGTGTCTGCAGACTCCAGGTTTCTCTCAGATCCAGTCCAGGTTGGCTCAGATACAG 162
Qy 7711 CAAGACAGTTGCAAGGGTTTCCAACTCAGGATCAGGATCAGGATCAGGATCAGGATCAGGAT 7770
Db 161 CAAGACAGTTGCAAGGGTTTCCAACTCAGGATCAGGATCAGGATCAGGATCAGGATCAGGAT 102
Qy 7771 CAGATAGAAATTA-ACGTGAAACACACCTCCAACTTCTAATCAAAAGTTGAAATCATTTCA 7829
Db 101 CAGATAGAAATTAAGGCGTGAACACACCTCCAACTTCTAATCAAAAG-TGAAATCATTTCA 43
Qy 7830 GAAACAGTGGTGTATGAAGCAATAATGCTGTAATAGAACATTT 7871
Db 42 GAAACAGTGGTGTATGAAGCAATAATGCTGTAATAGAACATTT 1

RESULT 12

US-10-033-528-1410/c
; Sequence 1410, Application US/10033528
; Publication No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1410
; LENGTH: 341
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1410

Query Match 3.2%; Score 316.4; DB 13; Length 341;
Best Local Similarity 99.1%; Pred. No. 4.6e-62;

Qy	1503	AGAAATTTTGGATCCATAAGAGCCAAAAAGGAGGACATTTGATAATGTTTAAAAAGCCCCAGA	1562
Db	1	AGAAATTTTGGATCCATAAGAGCCAAAAAGGAGGACATTTGATAATGTTTAAAAAGCCCCAGA	60
Qy	1563	AGAAACAGAAAAAGACAAAGATGAGACTGAGATGACTCTTAAAGATGCTCGAGAAAAACAG	1622
Db	61	AGAAACAGAAAAAGACAAAGATGAGACTGAGATGACTCTTAAAGATGCTCGAGAAAAACAG	120
Qy	1623	AGAAAGATTTGAAGACCCAGTCCCTTGAAAAAGACAGTGAGCAGACAAAAACACCATGATGA	1682
Db	121	AGAAAGATTTGAAGACCCAGTCCCTTGAAAAAGACAGTGAGCAGACAAAAACACCATGATGA	180
Qy	1683	CCCTGAGCAGAGAAAAATCTCGAGGAGCCACAGAAAGTTGGGGATAAAGGTAACTCTGTGTC	1742
Db	181	CCCTGAGCAGAGAAAAATCTCGAGGAGCCACAGAAAGTTGGGGATAAAGGTAACTCTGTGTC	240
Qy	1743	AGCAAAATCTGGCGACAAACACAACAATGCAACTTCAGAGAGACTAGTCCCTCTGAAGG	1802
Db	241	AGCAAAATCTGGCGACAAACACAACAATGCAACTTCAGAGAGACTAGTCCCTCTGAAGG	300
Qy	1803	GAGGAGCCCTGCTGGGGTGTCTCTCAGAAAAACCCCGATAGCAGCAACATGGCAGAGAAGAA	1862
Db	301	GAGGAGCCCTGCTGGGGTGTCTCTCAGAAAAACCCCGATAGCAGCAACATGGCAGAGAAGAA	360
Qy	1863	GGTGGCATCTGAGCTCCCGCAGGATGGCCAGAAACCTTAACAAGACATGTGTGAG	1917
Db	361	GGTGGCATCTGAGCTCCCGCAGGATGGCCAGAAACCTTAACAAGACATGTGTGAG	415
RESULT 8			
US-09-833-790-199			
; Sequence 199, Application US/09833790			
; Patent No. US20020068288A1			
; GENERAL INFORMATION:			
; APPLICANT: Lodes, Michael J.			
; APPLICANT: Wang, Tongtong			
; APPLICANT: Secretist, Heather			
; APPLICANT: Mohamath, Raodoh			
; APPLICANT: Indirias, Carol Y.			
; APPLICANT: Fan, Liqun			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY			
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER			
; FILE REFERENCE: 210121.512			
; CURRENT APPLICATION NUMBER: US/09/833,790			
; CURRENT FILING DATE: 2001-04-11			
; NUMBER OF SEQ ID NOS: 440			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 199			
; LENGTH: 574			
; TYPE: DNA			
; ORGANISM: Homo sapien			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (1)...(574)			
; OTHER INFORMATION: n = A,T,C or G			
US-09-833-790-199			
Query Match 3.6%; Score 359; DB 9; Length 574;			
Best Local Similarity 75.7%; Pred. No. 1.2e-71;			
Matches 571; Conservative 0; Mismatches 1; Indels 182; Gaps 3;			
Qy	2818	GAGAGAAACAGCAGNAGAAAGAACGATGCACCAAGCCGACATGGGTAAATACACATTT	2877
Db	1	GAGAGAAACAGCAGNAGAAAGAACGATGCACCAAGCCGACATGGGTAAATACACATTT	60
Qy	2878	CCAGTTAAGCATCAGGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTGCAGGATATGGT	2937
Db	61	CCAGTTAAGCATCAGGTTTGGAAACAAAAAGGTGAAGAGTACAGAGTGCAGGATATGGT	120
Qy	2938	GGTTGGAGCTGGATTTAGTAAACTCATGTTTATAGTTTGTTCCTTAATTTGCCAGCAAT	2997
Db	121	GGTTGGAGCTGGATTTAGTAAACTCATGTTTATAGTTTGTTCCTTAATTTGCCAGCAAT	180

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21036
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(516)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-21036

Query Match 4.68; Score 452.4; DB 11; Length 516;
Best Local Similarity 98.1%; Pred. No. 4.3e-93;
Matches 456; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2132 CTGGAGCTGGAAAGGAGCATCTGGCTCAACTCGAATCATCACAGATTGGGAATCCAG 2191
Db |||||
Qy 52 CTGGAGCTGGAAAGGAGCATCTGGCTCAACTCGAATCATCACAGATTGGGAATCCAG 111
Db |||||
Qy 2192 ATAGCAAACTAGTCAGCTGAAGAGCCAGGAGGCGCTGACATCAAGCAAAATA 2251
Db |||||
Qy 112 ATAGCAAACTAGTCAGCTGAAGAGCCAGGAGGCGCTGACATCAAGCAAAATA 171
Db |||||
Qy 2252 AATTATTAGAGGGGCAAGAGTACTGTAGTTAACTCTCAAGGAGAAATTTACGGT 2311
Db |||||
Qy 172 AATTATTAGAGGGGCAAGAGTACTGTAGTTAACTCTCAAGGAGAAATTTACGGT 231
Db |||||
Qy 2312 TGAGCACCACCAAGGAGTATCATGAAGGAATATCAAAATTTAAATTTGGTC 2371
Db |||||
Qy 232 TGAGCACCACCAAGGAGTATCATGAAGGAATATCAAAATTTAAATTTGGTC 291
Db |||||
Qy 2372 AAGAAGGGAAGTATCGCTTACCACAAATCAATCACTCCACCAATTCATTTGCTTGAATA 2431
Db |||||
Qy 292 AAGAAGGGAAGTATCGCTTACCACAAATCAATCACTCCACCAATTCATTTGCTTGAATA 351
Db |||||
Qy 2432 AGCACCAGCAGACAGAACCATGATGAAGAGGCAATCTTGCACATAAGTTCTGTCTGA 2491
Db |||||
Qy 352 AGCACCAGCAGACAGAACCATGATGAAGAGGCAATCTTGCACATAAGTTCTGTCTGA 411
Db |||||
Qy 2492 CTCACGAGGAGAGTTCAAAATGGAACGGTCTGTCCATGGTCCAAAGTTCTTACCATAT 2551
Db |||||
Qy 412 CTCACGAGGAGAGTTCAAAATGGAACGGTCTGTCCATGGTCCAAAGTTCTTACCATAT 471
Db |||||
Qy 2552 CTACTCTGAGACTGACTATCACCAATTAGAAACCAACATCCCTT 2596
Db |||||
Qy 472 CTACTCTGAGACTGACTATCACCAATTAGAAACCAACATCCCTT 516
Db |||||

RESULT 6
US-09-918-995-15144
; Sequence 15144, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15144
; LENGTH: 490

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(490)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-15144

Query Match 4.3%; Score 421; DB 11; Length 490;
Best Local Similarity 95.1%; Pred. No. 6.4e-86;
Matches 444; Conservative 0; Mismatches 22; Indels 1; Gaps 1;

Qy 9238 GTTATTGAAAAAAGAAAGAAAGCAAGAAAAAGATATATGGGTCAAGTGA 9297
Db |||||
Qy 25 GTTTTGAGACATTAGGCCAGNGCGNGGAATTCAAAAAGATATATGGGTCAAGTGA 84
Db |||||
Qy 9298 ACTCCATGGAATGCCAGCTCTCTTTCAGTGAAGAGCTGGTTAGAGTCTCACAA 9357
Db |||||
Qy 85 ACTCCATGGAATGCCAGCTCTCTTTCAGTGAAGAGCTGGTTAGAGTCTCACAA 144
Db |||||
Qy 9358 AACCTTTGACTGTATTTATTTATTTGTCAAAAAGAGCGCTTTTATTTGCTGCCTCAT 9417
Db |||||
Qy 145 AACCTTTGACTGTATTTATTTATTTGTCAAAAAGAGCGCTTTTATTTGCTGCCTCAT 204
Db |||||
Qy 9418 TTGTCAGTAAAGTATTTTCTTATAAAATCCAGCCCGGTTACATATAATCATCTGAT 9477
Db |||||
Qy 205 TTGTCAGTAAAGTATTTTCTTATAAAATCCAGCCCGGTTACATATAATCATCTGAT 264
Db |||||
Qy 9478 CTTATCATGATCTCTGTAGTAAAGTACAGAGACCTCTAGATGCTTTTCTTCTAT 9537
Db |||||
Qy 265 CTTATCATGATCTCTGTAGTAAAGTACAGAGACCTCTAGATGCTTTTCTTCTAT 324
Db |||||
Qy 9538 GAAAGGAGCTCTATGTACACATGTGCACACACACAACTGGGAATCAACAATGAGTTT 9597
Db |||||
Qy 325 GAAAGGAGCTCTATGTACACATGTGCACACACACAACTGGGAATCAACAATGAGTTT 384
Db |||||
Qy 9598 ATTGTTCAATGATGATTAATTAAGCTTGCATAAAGGTTGGGCTAAAGTGGCTTGGG 9657
Db |||||
Qy 385 ATTGTTCAATGATGATTAATTAAGCTTGCATAAAGGTTGGGCTAAAGTGGCTTGGG 443
Db |||||
Qy 9658 TACAGACTCTGTCCTTGAATATACAGTACAAATTTGTCATTAAT 9704
Db |||||
Qy 444 TACAGACTCTGTCCTTGAATATACAGTACAAATTTGTCATTAAT 490
Db |||||

RESULT 7
US-09-918-995-34912
; Sequence 34912, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34912
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(416)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-34912

Query Match 4.2%; Score 411.8; DB 11; Length 416;
Best Local Similarity 99.5%; Pred. No. 7.3e-84;
Matches 413; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20812
; LENGTH: 1148
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20812

Query Match 5.6%; Score 553.8; DB 12; Length 1148;
Best Local Similarity 89.8%; Pred. No. 4.4e-116;
Matches 628; Conservative 0; Mismatches 17; Indels 54; Gaps 1;

QY 2079 GATAGTAGGTGATTTCCATCGGAGAGTCCACGGGGAGCTAAGTGAATCTCCTGGAGC 2138
Db |||||
QY 430 GACAGTAGGTGATTTCAAATCGGAGAGTCCAAACGGGGAGCTAAGTGAATCTCCTGGAGC 489
Db |||||
QY 2139 TGAAGAGAGAGCATCTGGCTCAACTCGAATCATCACAGATTCGCGAATCCAGATAGCAA 2198
Db |||||
QY 490 TGAAGAGAGAGCATCTGGCTCAACTCGAATCATCACAGATTCGCGAATCCAGATAGCAA 549
Db |||||
QY 2199 ACTTAGTCAGCTCAAGAGCAGCAGGTGGCAGCGCTGCACATGAAGCAAAATAAATATT 2258
Db |||||
QY 550 ACTTAGTCAGCTCAAGAGCAGCAGGTGGCAGCGCTGCACATGAAGCAAAATAAATATT 609
Db |||||
QY 2259 TAAGGAGGCGCAAGAGTACTGGTAGTTAACTCTCAAGGAGAAATTTACGGTTGAGCAC 2318
Db |||||
QY 610 TAAGGAGGCGCAAGAGTACTGGTAGTTAACTCTCAAGGAGAAATTTACGGTTGAGCAC 669
Db |||||
QY 2319 CAAAAGGAGATGATCATGAAAGGAATATCAACAATTTATTTAAATTTGGTCAAGAAG 2378
Db |||||
QY 670 CGAAAAGGAGTGGTTCATGAAGAGAGTATCAACAATTTATTTAAATTTGGTCAAGAAG 729
Db |||||
QY 2379 GAAGTATCGGCTCTACCAAACTCAATCTCCACCAATTCATTTGCTTTGAATAAGCACCA 2438
Db |||||
QY 730 GAAGTATCGGCTCTACCAAACTCAATCTCCACCAATTCATTTGCTTTGAATAAGCACCA 789
Db |||||
QY 2439 GCACAGAGAGACCATGATAAGAGAGGATCTTGACATTAAGTTCTGTCTGACTCCAGC 2498
Db |||||
QY 790 GCACAGAGAGACCATGATAAGAGAGGATCTTGACATTAAGTTCTGTCTGACTCCAGC 849
Db |||||
QY 2499 AGCAGAGTCAAAATGGAAGCGTCTGTCCATGGGTCCAAAGTTCTTACCATATCTACTCT 2558
Db |||||
QY 850 AGCAGAGTCAAAATGGAAGCGTCTGTCCATGGGTCCAAAGTTCTTACCATATCTACTCT 909
Db |||||
QY 2559 GAGACTGACTATCACCACAAATAGAAAAACAACATCCCTTCATCCTTTTTTCATCCCACTG 2618
Db |||||
QY 910 GAGACTGACTATCACCACAAATAG- - - - - 932
Db |||||
QY 2619 GGCATCACATAGGGCAATTTGGATCAAGGAGTTCAGATGTGTAGCAAAACCCAGAGAAAT 2678
Db |||||
QY 933 - - - - - ACTGGATCAAGGAGTTCAGATGTGTAGCAAACTCAGAGAAAT 975
Db |||||
QY 2679 TGCATTTGGCTTTAGCCATTTTGGAGTGTGCAGTTAAACCAAGTTGTGATGCTACCAATATG 2738
Db |||||
QY 976 TGCATTTGGCTTTTGGAGTGTGCAGTTAAACCAAGTTGTGATGCTACCAATATG 1035
Db |||||
QY 2739 GCGAGAAATTTTGGAGCATACAGGTTTACACCGGATGAC 2777
Db |||||
QY 1036 GCGAGAAATCTTTAGGACATACAGCTTTCTCCCTTTGTC 1074

RESULT 4

US-09-960-706-151/c
; Sequence 151, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperp
; FILE OF INVENTION: Gene Expression Profiles
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 AA236286
US-09-960-706-151

Query Match 5.2%; Score 516.4; DB 12; Length 519;
Best Local Similarity 99.8%; Pred. No. 9.3e-108;
Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8750 AACCTATGGACCTTGGCCACCATGGAAGAGAGTACAAGACGATATTTATGAAAGCTGA 8809
Db |||||
QY 518 AACCTATGGACCTTGGCCACCATGGAAGAGAGTACAAGACGATATTTATGAAAGCTGA 459
Db |||||
QY 8810 CGAAATTTGGCGAGATATGACCAAAATTTTCTAATGCTCTTACTACAAATCCAAAGTG 8869
Db |||||
QY 458 CGAAATTTGGCGAGATATGACCAAAATTTTCTAATGCTCTTACTACAAATCCAAAGTG 399
Db |||||
QY 8870 ACTCCCATTTTACCAGTGTGCAGAGTTCTCGAATCATTTTGTACAGAAATTTGAAAG 8929
Db |||||
QY 398 ACTCCCATTTTACCAGTGTGCAGAGTTCTCGAATCATTTTGTACAGAAATTTGAAAG 339
Db |||||
QY 8930 GCTTCAAAGTACAGGTCTCATACAAACAACTGAGTCTAGCTTCTTAAAGTTTAAAG 8989
Db |||||
QY 338 GCTTCAAAGTACAGGTCTCATACAAACAACTGAGTCTAGCTTCTTAAAGTTTAAAG 279
Db |||||
QY 8990 CGTGTAAACCTAAACATAAAACACAGCAAGAACTCTGGTTGCTGCAACTATTTTAAATTAAG 9049
Db |||||
QY 278 CGTGTAAACCTAAACATAAAACACAGCAAGAACTCTGGTTGCTGCAACTATTTTAAATTAAG 219
Db |||||
QY 9050 GAGCCAGATGTTTTTGTAGTCAGGCTATCCTGACAGAGTCTGACCTTAAACTTCGTTTTTATT 9109
Db |||||
QY 218 GAGCCAGATGTTTTTGTAGTCAGGCTATCCTGACAGAGTCTGACCTTAAACTTCGTTTTTATT 159
Db |||||
QY 9110 GGTCAATAACAGTCCAATATATTTCTTGGCCAAATTTTGTCCACGCGACAGAAAAAGCAA 9169
Db |||||
QY 158 GGTCAATAACAGTCCAATATATTTCTTGGCCAAATTTTGTCCACGCGACAGAAAAAGCAA 99
Db |||||
QY 9170 AGTCAACGACACCATATATCTTGTCAAGATCAGATGTTTTTACTATTTGTGCGAGAGCGAG 9229
Db |||||
QY 98 AGTCAACGACACCATATCTTGTCAAGATCAGATGTTTTTACTATTTGTGCGAGAGCGAG 39
Db |||||
QY 9230 AAAACCTTTGTTTTTGTAAAAAAGAAAAAGAAAGCA 9267
Db |||||
QY 38 AAAACCTTTGTTTTTGTAAAAAAGAAAAAGAAAGCA 1

RESULT 5

US-09-918-995-21036
; Sequence 21036, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.

8564 ATGAGTATCTCTCCACAGTCCAGTCAACAGAGATGCCATGACAGTGTCTCACGCCAC 8623
|||||
299 ATGAGTATCTCTCCACAGTCCAGTCAACAGAGATGCCATGACAGTGTCTCACGCCAC 358
|||||
8624 TTAACAGAGAAGGATTATGAGGGGTTGAAGAGGGTCTCCGTTCCCTTACAGGCCCAATAAGA 8683
|||||
359 TTAACAGAGAAGGATTATGAGGGGTTGAAGAGGGTCTCCGTTCCCTTACAGGCCCAATAAGA 418
|||||
8684 TGGCCCTGGCCCTTCCCTTGAACAGATAGACCCCTAAATGATGACACAGATTAATATGPGTTA 8743
|||||
419 TGGCCCTGGCCCTTCCCTTGAACAGATAGACCCCTAAATGATGACACAGATTAATATGPGTTA 478
|||||
8744 TTAAGGAACCTTATGAGCCTTGGCCACCATGGAAGAAAGATCAAAAGACGATATTATGAAA 8803
|||||
479 TTAAGGAACCTTATGAGCCTTGGCCACCATGGAAGAAAGATCAAAAGACGATATTATGAAA 538
|||||
8804 AGCTGACGGAATTTGTCGAGATATGACCAAAATTTTGTAACTCTCGTTACTACAAATC 8863
|||||
539 AGCTGACGGAATTTGTCGAGATATGACCAAAATTTTGTAACTCTCGTTACTACAAATC 598
|||||
8864 CAAGTGACCTCCCATTTTACCAGTGTGACAGAGTCTCGAATCATCTTTGTACAGAAAT 8923
|||||
599 CAAGTGACCTCCCATTTTACCAGTGTGACAGAGTCTCGAATCATCTTTGTACAGAAAT 658
|||||
8924 TGAAGGCTTCAAGCTACAGGTCTCATAAACAACAACTGCAGTCTACAGCTTCTTAA 8983
|||||
659 TGAAGGCTTCAAGCTACAGGTCTCATAAACAACAACTGCAGTCTACAGCTTCTTAA 718
|||||
8984 GTTCAGCGTGTAACTCAATTAACACAGCAAGAACTCTGGTGTCTGAACCTATTATTTAA 9043
|||||
719 GTTCAGCGTGTAACTCAATTAACACAGCAAGAACTCTGGTGTCTGAACCTATTATTTAA 778
|||||
9044 ATTAAGAGCCAGATGTTTTTGTAGTCAGGCTATCCTGACAAGACTTTCAGCTTAACTTCGTT 9103
|||||
779 ATTAAGAGCCAGATGTTTTTGTAGTCAGGCTATCCTGACAAGACTTTCAGCTTAACTTCGTT 838
|||||
9104 TTATTTGGTCATAACAGTCCAATTAATTTCTGGCCAAATTTTGTCCAAGCGCAAGAAAA 9163
|||||
839 TTATTTGGTCATAACAGTCCAATTAATTTCTGGCCAAATTTTGTCCAAGCGCAAGAAAA 898
|||||
9164 AAGCAAGTCAACGACACCATTTCTGTCAAGATCAGATGTTTTTACTATTGTGGCAGA 9223
|||||
899 AAGCAAGTCAACGACACCATTTCTGTCAAGATCAGATGTTTTTACTATTGTGGCAGA 958
|||||
9224 ACGGAGAAACTTTGTTTTTGAAGAAAAAGAAAGAAAGCAAGAAAAAGATA 9280
|||||
959 ACGGAGAAACTTTGTTTTTGAAGAAAAAGAAAGAAAGCAAGAAAAAGATA 1015
|||||

RESULT 2

US-09-925-297-357
; Sequence 357, Application US/09925297
; Patent No. US20020081659A1

GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 357

; LENGTH: 1408
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1396)

; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-357

Query Match 6.8%; Score 673.8; DB 9; Length 1408;
Best Local Similarity 97.1%; Pred. No. 1.6e-143;
Matches 675; Conservative 9; Mismatches 11; Indels 0; Gaps 0;
QY 2083 GTAGGTGATTTTCATCTCGGAGAGTCCACGGGGAGCTAAGTGAATCTCTCGAGCTGGA 2142
Db 276 GTAGGTGATTTCAATTCGGAGAGTCCACGGGGAGCTAAGTGAATCTCTCGAGCTGGA 335
QY 2143 AAAGGAGCATCTGGCTCAACTCGAATCATCACCAGATTCGGGAATCCAGATACCAACTT 2202
Db 336 AAAGGAGCATCTGGCTCAACTCGAATCATCACCAGATTCGGGAATCCAGATACCAACTT 395
QY 2203 AGTCAGCTGAAGAGCAGCAGGTGGCAGCGCTGCACATGAAGCAAAATAAATTTATTAAG 2262
Db 396 AGTCAGCTGAAGAGCAGCAGGTGGCAGCGCTGCACATGAAGCAAAATAAATTTATTAAG 455
QY 2263 GAGGGCAAGAGGTACTGTAGTTAACTCTCAAGGAGAAATTTACGGTTGACGACCAA 2322
Db 456 GAGGGCAAGAGGTACTGTAGTTAACTCTCAAGGAGAAATTTACGGTTGACGACCAA 515
QY 2323 AAAGGAGTGTATGATGAAGAGAAATATCAACAATTAATTTAAATTTGGTCAAGAGGAAG 2382
Db 516 AAAGGAGTGTATGATGAAGAGAAATATCAACAATTAATTTAAATTTGGTCAAGAGGAAG 575
QY 2383 TATCGGCTTACCAATCAATTAATTTCAATTTGCTTTGAATTAAGCACCAGCAC 2442
Db 576 TATCGGCTTACCAATCAATTAATTTCAATTTGCTTTGAATTAAGCACCAGCAC 635
QY 2443 AGAAGAGCATTGATGAAGAGAGGCAATCTTGACATTAAGTTCTGTCTGACTCAGCAGGA 2502
Db 636 AGAAGAGCATTGATGAAGAGAGGCAATCTTGACATTAAGTTCTGTCTGACTCAGCAGGA 695
QY 2503 GAGTTCAATGAAGAGGTTCTGTCCATGGTCCAAAGTTCTTACCATACTACTCTGAGA 2562
Db 696 GAGTTCAATGAAGAGGTTCTGTCCATGGTCCAAAGTTCTTACCATACTACTCTGAGA 755
QY 2563 CTGACTATCAACCAATTAGAAGCAACATCCCTTCATCTCTTTTCCATCCCACTGGCA 2622
Db 756 CTGACTATCAACCAATTAGAAGCAACATCCCTTCATCTCTTTTCCATCCCACTGGCA 815
QY 2623 TCATATAGGCAAAATTTGATCAAGGAGTTGATGTTAGCAAAAGTTGATGTTAGCAAA 2682
Db 816 TCATATAGGCAAAATTTGATCAAGGAGTTGATGTTAGCAAAAGTTGATGTTAGCAAA 875
QY 2683 TTGGCTTTAGCCATTTTGGAGTGTGAGTTTAAACAGTTGATGTTACCAATATGCGCA 2742
Db 876 TTGGCTTTAGCCATTTTGGAGTGTGAGTTTAAACAGTTGATGTTACCAATATGCGCA 935
QY 2743 GAATTTTATAGGACATACAGGTTTACACCGGATGAC 2777
Db 936 GAATTTTATAGGACATACAGGTTTCTCCCTTTGTC 970

RESULT 3

US-09-814-353-20812
; Sequence 20812, Application US/09814353
; Publication No. US20030165831A1

GENERAL INFORMATION:

; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124

Result No.	Score	Query Match	Length	DB	ID	Description
1	814.6	8.3	1032	9	US-09-729-855-21	Sequence 21, Appl
2	673.8	6.8	1408	9	US-09-925-297-357	Sequence 357, App
3	553.8	5.6	1148	12	US-09-814-353-20812	Sequence 20812, A
C 4	516.4	5.2	519	12	US-09-960-706-151	Sequence 151, App
5	452.4	4.6	516	11	US-09-918-995-21036	Sequence 21036, A
6	421	4.1	490	11	US-09-918-995-15144	Sequence 15144, A
7	411.8	4.2	416	11	US-09-918-995-34912	Sequence 34912, A
8	359	3.6	574	9	US-09-833-790-199	Sequence 199, App
9	327.8	3.3	477	11	US-09-918-995-3226	Sequence 3226, App
C 10	316.4	3.2	341	10	US-09-920-3008-1410	Sequence 1410, App
C 11	316.4	3.2	341	12	US-09-929-826-1410	Sequence 1410, App
C 12	316.4	3.2	341	13	US-10-033-828-1410	Sequence 8224, App
13	223.2	2.3	545	10	US-09-796-692-8224	Sequence 8224, App
14	223.2	2.3	545	14	US-10-040-862-8224	Sequence 862, App
15	181.4	1.8	405	10	US-09-728-445-869	Sequence 281, App
C 16	177.4	1.8	2666	11	US-09-822-846-281	Sequence 846, App

Qy 2091 TTAAGTGEORQ-----SKLSPQMGVHODKTLPPA-----OSSSVGPAPAKAPQTAQP----- 2136
 Db 1955 RQIASDKDARERGSSDSSSLSHRYETASDAIEVISPASSPAPQPKQEPQAYQPDWVK 2014
 Qy 2137 -----SARPOPTQPSQAPQEVQTO-----PEVOTQTIVTSSHVP 2171
 Db 2015 ANQAEENSTRQYEGPLHHYRQESPSPPQPPPLPSSQSEGMGQVPRTHRLTLADHIC 2074
 Qy 2172 SEAPTHAQSCKPOVAAGSQPSNVQSGSPVRVQSPQSTQIRPSTPSQLSPGQSQVQTT 2231
 Db 2075 QIITQDFARQVFSQASTSTFQT-----SP-----SAUSSTPVRTKTSRYSPESQSQ---T 2123
 Qy 2232 TSQPIP---IQPHTSLQIPSGQPSQPSQPSQSTQTLSSQTLNVQSVSPSPRQLOIQ 2288
 Db 2124 VLHPRCPGPRVSPENLVKSRGSRPKSKPE-----RSHIPSPPEPISP 2166
 Qy 2289 PQPOVIAPQLOQQVQLSQ 2308
 Db 2167 PQGP--AVHEKQDSMLLSQ 2184

RESULT 15
 AKAP_HUMAN
 ID AKAP_HUMAN STANDARD; PRT; 3911 AA.
 AC Q99996; O14869; O43355; O94895; Q9UQH3; Q9UQQ4; Q9Y6B8; Q9Y6Y2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update).
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE A-kinase anchor protein 9 (Protein Kinase A anchoring protein 9)
 DE (PKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
 DE protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
 DE (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized
 DE PKN-associated protein) (CG-NAP).
 GN AKAP9 OR AKAP450 OR AKAP350 OR KIA0803.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Brain;
 RX MEDLINE=98151389; PubMed=9482789;
 RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;
 RT "Yotiao, a novel protein of neuromuscular junction and brain that
 RT interacts with specific splice variants of NMDA receptor subunit
 RT NR1.1".
 RL J. Neurosci. 18:2017-2027(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
 RX MEDLINE=99219864; PubMed=10202149;
 RA Witczak O., Skaalhegg B.S., Keryer G., Bornens M., Tasken K.,
 RA Jahnson T., Oerstavik S.;
 RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring
 RT protein located in the centrosome, AKAP450.";
 RL EMBO J. 18:1858-1868(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;
 RX MEDLINE=99287934; PubMed=10358086;
 RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;
 RT "Characterization of a novel giant scaffolding protein, CG-NAP, that
 RT anchors multiple signaling enzymes to centrosome and the Golgi
 RT apparatus.";
 RL J. Biol. Chem. 274:17267-17274(1999).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Kemmer W.A., Deiss S., Schwarz U.;
 RT "Cloning of Hyperion.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
 RC TISSUE=Gastric parietal cell;

RX MEDLINE=99115654; PubMed=9915845;
 RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
 RA Trotter K.W., Milgram S.L., Goldenring J.R.;
 RT "AKAP350, a multiply spliced protein kinase A-anchoring protein
 RT associated with centrosomes.";
 RL J. Biol. Chem. 274:3055-3066(1999).
 RN [6]
 RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
 RC TISSUE=Lymphoblast;
 RA Hinds K., Sutterer C., Becker M., Hawkins M.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
 RC TISSUE=Lung;
 RA Milgram S.L., Goldenring J.R., Schmidt P.H.;
 RT "AKAP350: A multiply spliced family of proteins with centrosomal
 RT association.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
 RC TISSUE=Brain;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RL for large proteins in vitro.";
 RN DNA Res. 5:277-286(1998).
 RN [9]
 RP SEQUENCE OF 17-1800 FROM N.A.
 RA Wu X., Graves T., Bradshaw H.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
 CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASE AND
 CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL
 CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN
 CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-
 CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR
 CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS
 CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.
 CC -I- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N
 CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)
 CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.
 CC -I- SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND
 CC CYTOPLASMIC IN PARIETAL CELLS.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=6;
 CC Name=1;
 CC IsoId=Q99996-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q99996-2; Sequence=VSP_004102, VSP_004107;
 CC Name=3; Synonyms=CG-NAP;
 CC IsoId=Q99996-3; Sequence=VSP_004102, VSP_004105, VSP_004107;
 CC Name=4; Synonyms=Yotiao;
 CC IsoId=Q99996-4; Sequence=VSP_004103, VSP_004104;
 CC Name=5;
 CC IsoId=Q99996-5; Sequence=VSP_004108;
 CC Name=6; Synonyms=AKAP350;
 CC IsoId=Q99996-6; Sequence=VSP_004106, VSP_004107, VSP_004109;
 CC -I- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY
 CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.
 CC -I- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
 CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
 CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
 CC -I- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
 CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.
 CC -I- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR
 CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

THIS PAGE BLANK (USPTO)

